

SPRINGER NATURE
Reference



Cheng-Few Lee
Alice C. Lee
Editors

Encyclopedia of Finance

Third Edition

 Springer

Encyclopedia of Finance

Cheng-Few Lee • Alice C. Lee
Editors

Encyclopedia of Finance

Third Edition

With 194 Figures and 312 Tables

 Springer

Editors

Cheng-Few Lee
Department of Finance & Economics
Rutgers University
Piscataway, NJ, USA

Alice C. Lee
Center for PBBEF Research
Phoenix, AZ, USA

ISBN 978-3-030-91230-7

ISBN 978-3-030-91231-4 (eBook)

<https://doi.org/10.1007/978-3-030-91231-4>

1st edition: © Springer Science+ Business Media, Inc. 2006

2nd edition: © Springer Science+Business Media New York 2013

3rd edition: © Springer Nature Switzerland AG 2022

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface to the Third Edition

Since the first edition was published in 2006 and the second edition was published in 2013, this encyclopedia has been very popular in both academic and practitioner professions. It has been one of the most downloaded books, which was published by Springer, in the area of finance and economics.

In this new edition, we have revised Part I, Part II, and the appendices extensively. In Part I, we added more than 200 terminologies and essays. In Part II, we added 46 new chapters. Finally, we added four new appendices. The new chapters and appendices can be found in the table of contents.

One hundred and nine papers included in Part II can be classified into ten groups as follows:

- (a) Investment analysis and portfolio management (Chaps. 4, 8, 11, 13, 20, 22, 30, 32, 35, 41, 46, 49, 55, 62, 66, 84, 97, 98, 101, 105, 107, and 109)
- (b) Financial management and corporate finance (Chaps. 12, 19, 23, 27, 28, 29, 33, 40, 43, 52, 57, 68, 69, 70, 71, 72, 73, 79, 86, 87, 88, 90, 91, 99, 102, 103, 104, and 106)
- (c) International finance (Chaps. 5, 7, 16, 31, 34, 42, 43, 48, 51, 53, and 67)
- (d) Microstructure (Chaps. 17, 18, 21, 31, 36, 37, 38, 39, and 45)
- (e) Asset pricing (Chaps. 9, 10, 11, 13, 35, 58, and 63)
- (f) Financial institutions and markets (Chaps. 2, 3, 14, 25, 47, 54, 77)
- (g) Derivatives (Chaps. 6, 29, 44, and 65)
- (h) Real estate finance (Chaps. 15, 26, 59, 60, and 96)
- (i) Risk management (Chaps. 5, 6, 7, 23, 24, 25, 40, 56, 60, 61, 64, 74, 75, 76, 81, 82, 83, 85, 94, and 95)
- (j) Methodology and fintech (Chaps. 78, 80, 89, 92, 93, 100, and 108)

For both undergraduate and graduate students, this encyclopedia is a good supplementary material for the above-listed finance courses. In addition, this encyclopedia can be a good supplementary material for financial accounting courses. We believe that this encyclopedia will not only be useful to students but also for professors and practitioners in the field of finance as reference.

We would like to thank the contributors for their willingness to share their expertise and their thoughtful essays in Part II. We would like to thank Andrew

Spencer and Nitza Jones of Springer for their coordination and suggestions for this book. Finally, we would also like to express our gratitude to our secretary and assistant, Sophia Chen, Wen-Chie Yeh, Natalie Krawczyk, and Matthew Gregory for their efforts in helping us pull together this tremendous repository of information.

We hope that the readers will find the encyclopedia to be an invaluable resource.

NJ, USA
AZ, USA
June 2022

Cheng-Few Lee
Alice C. Lee

Preface to the Second Edition

Since the first edition was published in 2006, this encyclopedia has been very popular in both academic and practitioner professions. It has been the most downloaded book in the area of finance and economics, which was published by Springer.

In this new edition, we have revised Part I, Part II, and Appendices extensively. In Part I, we added more than 200 terminologies and essays. In Part II, we added 24 new chapters. Finally, we added four new appendices. The new chapters and appendices can be found in the table content.

Seventy-four papers included in Part II can be classified into eight groups as follows:

- (a) Investment analysis and portfolio management (Chaps. 4, 8, 11, 13, 20, 22, 30, 32, 35, 41, 46, 49, 55, 62, and 66)
- (b) Financial management and corporate finance (Chaps. 12, 19, 23, 27, 28, 29, 33, 40, 43, 52, 57, 64, 68, 69, 70, 71, 72, and 73)
- (c) International finance (Chaps. 5, 7, 16, 31, 34, 42, 43, 48, 51, 53, and 67)
- (d) Microstructure (Chaps. 17, 18, 21, 31, 36, 37, 38, 39, and 45)
- (e) Asset pricing (Chaps. 9, 10, 11, 13, 35, 58, and 63)
- (f) Financial institutions and markets (Chaps. 2, 3, 14, 25, 47, and 54)
- (g) Derivatives (Chaps. 6, 29, 44, and 65)
- (h) Real estate finance (Chaps. 15, 26, 59, and 50)
- (i) Risk management (Chaps. 5, 6, 7, 23, 24, 25, 40, 56, 60, 61, 74, and 75)

For both undergraduate and graduate students, this encyclopedia is a good supplementary material for the above-listed finance courses. In addition, this encyclopedia can be a good supplementary material for financial accounting courses. We believe that this encyclopedia will not only be useful to students but also for professors and practitioners in the field of finance as a reference.

We would like to thank the contributors for willingness to share their expertise and their thoughtful essays in Part II. We would like to thank Mr. Brian J. Foster of Springer for his coordination and suggestions to this book. Finally, we would also

like to express our gratitude to our secretary and assistant, Ms. Miranda Mei-Lan Luo and Tzu Tai, for their efforts in helping us pull together this tremendous repository of information.

We hope that the readers will find the encyclopedia to be an invaluable resource.

NJ, USA

AZ, USA

Cheng-Few Lee

Alice C. Lee

Preface to the First Edition

Finance has become one of the most important and popular subjects in management school today. This subject has progressed tremendously in the last 40 years, integrating models and ideas from other areas such as physics, statistics, and accounting. The financial markets have also rapidly expanded and changed extensively because of improvement in technology and the ever-changing regulatory and social environment. For example, there has been a rapid expansion of financial concepts, instruments, and tools due to increased computing power and seemingly instantaneous information sharing through networks. The internationalization of businesses and economies will continue to impact the field of finance. With all this progress and expansion in finance and society, we thought that it would be useful to put together an updated comprehensive encyclopedia as a reference book for both students and professionals in an attempt to meet the demand for a key source of fundamental finance terminology and concepts.

This *Encyclopedia of Finance* contains five parts. Part I includes finance terminology and short essays. Part II includes 50 important finance chapters by well-known scholars and practitioners, such as James R. Barth, Ren-RAW Chen, Thomas C. Chiang, Quentin C. Chu, Wayne E. Ferson, Joseph E. Finnerty, Thomas S.Y. Ho, C.H. Ted Hong, Cheng Hsiao, Jing- Zhi Huang, Mao-wei Hung, John S Jahera Jr., Haim Levy, Wilbur G. Lewellen, Joseph P. Ogden, Fai-nan Peng, Gordon S. Roberts, Robert A. Schwartz, K.C. John Wei, and Gillian Yeo, among others. Topics covered in both Part I and Part II include fundamental subjects such as financial management, corporate finance, investment analysis and portfolio management, options and futures, financial institutions, international finance, and real estate finance. Part III contains appendices which discuss and derive some fundamental finance concepts and models, Part IV lists references, and Part V provides both subject and author indexes.

Fifty papers included in Part II can be classified into eight groups as follows:

- (a) Investment analysis and portfolio management (Chaps. 3, 7, 10, 12, 19, 21, 29, 31, 34, 40, 45, and 48)
- (b) Financial management and corporate finance (Chaps. 11, 18, 22, 26, 27, 28, 32, 39, and 42)
- (c) International finance (Chaps. 4, 6, 15, 30, 33, 41, 42, 47, and 50)

- (d) Microstructure (Chaps. 16, 17, 20, 30, 35, 36, 37, 38, and 44)
- (e) Asset pricing (Chaps. 8, 9, 10, 12, and 34)
- (f) Financial institutions and markets (Chaps. 1, 2, 13, 24, and 46)
- (g) Derivatives (Chaps. 5, 28, and 43)
- (h) Real estate finance (Chaps. 14, 25, and 49)
- (i) Risk management (Chaps. 4, 5, 6, 22, 23, 24, and 39)

For both undergraduate and graduate students, this encyclopedia is a good supplementary material for the above-listed finance courses. In addition, this encyclopedia can be a good supplementary material for financial accounting courses. We believe that this encyclopedia will not only be useful to students but also for professors and practitioners in the field of finance as a reference.

We would like to thank the contributors for willingness to share their expertise and their thoughtful essays in Part II. We would like to thank Ms. Judith L. Pforr, of Springer, for her coordination and suggestions to this book. Finally, we would also like to express our gratitude to our secretaries Ms. Miranda Mei-Lan Luo, Ms. Sue Wang, Ms. Ting Yen, and Ms. Meetu Zalani for their efforts in helping us pull together this tremendous repository of information.

We hope that the readers will find the encyclopedia to be an invaluable resource.

NJ, USA
AZ, USA

Cheng-Few Lee
Alice C. Lee

Contents

Volume 1

Part I	Terms and Essays	1
1	Terms and Essays	3
	Cheng-Few Lee and Alice C. Lee	
Part II	Papers	509
2	Deposit Insurance Schemes	511
	James R. Barth, Nguyen Nguyen, and Jiayi Xu	
3	Gramm-Leach-Bliley Act: Creating a New Bank for a New Millennium	529
	James R. Barth and Shen Zhang	
4	Pre-funded Coupon and Zero-Coupon Bonds: Cost of Capital Analysis	541
	Suresh Srivastava and Ken Hung	
5	Intertemporal Risk and Currency Risk	555
	Jow-Ran Chang and Mao-Wei Hung	
6	Credit Derivatives	573
	Ren-Raw Chen and Jing-Zhi Huang	
7	Foreign Exchange Risk Premium and Policy Uncertainty	585
	Thomas C. Chiang	
8	Treasury Inflation-Protected Securities	601
	Quentin C. Chu and Deborah N. Pittman	
9	Asset Pricing Models	613
	Wayne E. Ferson	
10	Conditional Asset Pricing	629
	Wayne E. Ferson	

11	Conditional Performance Evaluation	639
	Wayne E. Ferson	
12	Working Capital and Cash Flow	651
	Joseph E. Finnerty	
13	Evaluating Fund Performance Within the Stochastic Discount Factor Framework	667
	Jonathan Fletcher	
14	Duration Concepts, Analysis, and Applications	681
	Zvika Afik, Iraj Fooladi, Gady Jacoby, and Gordon Roberts	
15	Loan Contract Terms	703
	Aron Gottesman	
16	Chinese A and B Shares	713
	Yan He	
17	Decimal Trading in the U.S. Stock Markets	719
	Yan He	
18	The 1997 NASDAQ Trading Rules	723
	Yan He	
19	Reincorporation	729
	Randall A. Heron and Wilbur G. Lewellen	
20	Mean Variance Portfolio Allocation	743
	Cheng Hsiao and Shin-Huei Wang	
21	Online Trading	753
	Chang-Tseh Hsieh	
22	A Critical Evaluation of the Portfolio Performance Indices Under Rank Transformation	761
	Ken Hung, Chin W. Yang, Matthew Brigida, and Dwight B. Means	
23	Corporate Failure: Definitions, Methods, and Failure Prediction Models	773
	Jenifer Piesse, Cheng-Few Lee, Hsien-Chang Kuo, and Lin Lin	
24	Main Bank Relationships, Debt Structure, and Innovation in Japan	791
	Hai-Chin Yu and Phuong Dung Tong	
25	Term Structure: Interest Rate Models	819
	Thomas S. Y. Ho and Sang Bin Lee	
26	Review of REIT and MBS	835
	Chiuling Lu and Chou-Yen Wu	

27	Experimental Economics and the Theory of Finance	849
	Haim Levy	
28	Merger and Acquisition: Definitions, Motives, and Market Responses	877
	Jenifer Piesse, Cheng-Few Lee, Lin Lin, and Hsien-Chang Kuo	
29	Multistage Compound Real Options: Theory and Application	895
	William T. Lin, Cheng-Few Lee, and Chang-Wen Duan	
30	Market Efficiency Hypothesis	937
	Melody Lo	

Volume 2

31	The Microstructure/Micro-Finance Approach to Exchange Rates	945
	Melody Lo	
32	Arbitrage and Market Frictions	951
	Shashidhar Murthy	
33	Fundamental Tradeoffs in the Publicly Traded Corporation	963
	Joseph P. Ogden	
34	The Mexican Peso Crisis	973
	Fai-Nan Perng	
35	Methods for Portfolio Performance Evaluation	983
	Lalith P. Samarakoon and Tanweer Hasan	
36	Call Auction Trading	991
	Robert A. Schwartz and Reto Francioni	
37	Market Liquidity	1001
	Robert A. Schwartz and Lin Peng	
38	Market Makers	1007
	Robert A. Schwartz and Lin Peng	
39	Structure of Securities Markets	1011
	Robert A. Schwartz and Lin Peng	
40	Accounting Scandals and Implications for Directors: Lessons from Enron	1017
	Pearl Tan and Gillian Yeo	
41	Agent-Based Models of Financial Markets	1027
	Nicholas S. P. Tay	

42	The Asian Bond Market	1035
	Khairy Tourk and Hui Li	
43	Cross-Border Mergers and Acquisitions	1057
	Geraldo M. Vasconcellos and Richard J. Kish	
44	Jump Diffusion Model	1073
	Shin-Huei Wang	
45	Networks, Nodes, and Priority Rules	1093
	Daniel G. Weaver	
46	The Momentum Trading Strategy	1109
	K. C. John Wei and Linti Zhang	
47	Equilibrium Credit Rationing and Monetary Non-neutrality in a Small Open Economy	1123
	Ying Wu	
48	Policy Coordination Between Wages and Exchange Rates in Singapore	1137
	Ying Wu	
49	The Le Chatelier Principle of the Capital Market Equilibrium	1149
	Chin W. Yang, Ken Hung, Matthew Brigida, and John A. Fox	
50	MBS Valuation and Prepayments	1157
	C. H. Ted Hong and Wen-Ching Wang	
51	The Impacts of IMF Bailouts in International Debt Crises	1177
	Zhaohui Zhang and Khondkar E. Karim	
52	Corporate Governance: Structure and Consequences	1187
	Bikki Jaggi	
53	A Survey Article on International Banking	1223
	James Winder	
54	Hedge Funds: Overview, Strategies, and Trends	1249
	John M. Longo	
55	An Appraisal of Modeling Dimensions for Performance Appraisal of Global Mutual Funds	1273
	G. V. Satya Sekhar	
56	Structural Credit Risk Models: Endogenous Versus Exogenous Default	1293
	Michael B. Imerman	

57	Arbitrage Opportunity Set and the Role of Corporations	1317
	James S. Ang and Yingmei Cheng	
58	Equity Premium Puzzle: The Distributional Approach	1345
	Nadezhda Safronova	
59	Understanding Ginnie Mae Reverse Mortgage H-REMICs: Its Programs and Cashflow Analysis	1373
	C. H. Ted Hong and George H. Lee	
60	An Analysis of Risk Treatment in the Field of Finance	1397
	Fernando Gómez-Bezares and Fernando R. Gómez-Bezares	
61	The Trading Performance of Dynamic Hedging Models: Time Varying Covariance and Volatility Transmission Effects	1411
	Michael T. Chng and Gerard L. Gannon	
62	Portfolio Insurance Strategies	1437
	Lan-chih Ho, John Cadle, and Michael Theobald	
63	Time-Series and Cross-Sectional Tests of Asset Pricing Models	1467
	Kyung-Jin Choi, Dongcheol Kim, and Soon-Ho Kim	
64	Unified Model Arbitrage-Free Term Structure of Flow Risks . . .	1485
	Thomas S. Y. Ho and Sang Bin Lee	
65	A Comparison of Formulas to Compute Implied Standard Deviation	1511
	James S. Ang, Gwoduan David Jou, and Tsong-Yue Lai	
66	Securities Transaction Taxes: Literature and Key Issues	1531
	Anna Pomeranets	
67	Financial Control and Transfer Pricing	1541
	Savita A. Sahay	
68	Alternative Models for Evaluating Convertible Bond: Review and Integration	1563
	Cheng-Few Lee, Lie-Jane Kao, and Po-Cheng Wu	
69	A Rationale for Hiring Irrationally Overconfident Managers . . .	1581
	Oded Palmon and Itzhak Venezia	
70	Current Versus Permanent Earnings for Estimating Alternative Dividend Payment Behavioral Model: Theory, Methods and Applications	1599
	Cheng-Few Lee, Hong-Yi Chen, Alice C. Lee, and Yuhsin Tai	

71	Valuation of Interest Tax Shields	1641
	Michael Dothan	
72	Usefulness of Cash Flow Statements	1657
	Savita A. Sahay	
73	Do CEO Gender and Marital Status Affect Firm's R&D and Value? An Empirical Analysis Using Nonlinear Models	1685
	Keshab Shrestha, Cheng-Few Lee, and Abdul Ghafoor	
74	Three Alternative Methods for Estimating Hedge Ratios	1703
	Sheng-Syan Chen, Cheng-Few Lee, Fu-Lai Lin, and Keshab Shrestha	
75	Credit Risk Modeling: A General Framework	1727
	Ren-Raw Chen	
76	Bankruptcy Prediction Studies Across Countries Using Multiple Criteria Linear Programming (MCLP) and Other Data Mining Approaches	1765
	Wikil Kwak and Yong Shi	
77	Application of Difference-in-Differences Strategies in Finance: The Case of Natural Disasters and Bank Responses	1779
	James R. Barth, Kang Bok Lee, Xuan Shen, and Yeo Song Yoon	
78	Financial Panel Data Models, Strict Versus Contemporaneous Exogeneity, and Durbin-Wu-Hausman Specification Tests	1799
	Robert H. Patrick	
79	Accruals and the Asymmetric Timeliness of Earnings: A Decomposition Analysis	1829
	Wenhsin Hsu	
80	Computing Technology for Financial Service	1869
	Fang-Pang Lin	
81	Local Volatility Interest Rate Model	1901
	Thomas S. Y. Ho and Sang Bin Lee	
82	Applications of Logistic Regression and Hazard Method in Accounting and Finance Research	1919
	Feng Gao	
83	Cube Root Utility Theory	1939
	Jack Clark Francis	

Volume 3

84	A Global Comparative Study of Impact Investments Research in Academic Institutions	1963
	Rachel Calipha and Itzhak Venezia	
85	Financial Crisis, Capital Requirement, and Stress Tests: Evidence from the Extreme Value and Stable Paretian Estimates	1983
	Tony Sio-Chong U and Jacky Yuk-Chow So	
86	The Economics of and Accounting for Lease Transactions	2019
	Cheng-Few Lee and Ryan McDonough	
87	Pension Accounting, Inside Debt, and Capital Structure	2041
	Cheng-Few Lee and James Juichia Lin	
88	The Role of Earnings Management in Equity Valuation	2061
	Guanming He, April Zhichao Li, and Dongxiao Shen	
89	The Applications of Machine Learning in Accounting and Auditing Research	2095
	Hanxin Hu and Ting Sun	
90	Internal Capital Budgeting and Allocation in Financial Firms	2117
	Woo-Young Kang	
91	Job Security and CEO Compensation	2139
	James S. Ang and Wei Chen	
92	Tail-Risk Protection: Machine Learning Meets Modern Econometrics	2177
	Bruno Spilak and Wolfgang Karl Härdle	
93	Structural Breaks in Financial Panel Data	2213
	Yiannis Karavias	
94	More on Equilibrium Credit Rationing and Interest Rates: A Theory with New Evidence	2229
	Ying Wu	
95	The Effect of Basel III on Banks' Lending	2251
	Hao Chang and Yan Xue	
96	Mortgage Analysis	2263
	Aron Gottesman	

97	A History of Commercially Available Risk Models	2275
	John Blin, John Guerard, and Andrew Mark	
98	Short Selling Activity and Effects on Financial Markets and Corporate Decisions	2313
	Xu Guo and Chunchi Wu	
99	Simultaneous Equation Models for Financial Planning and Forecasting	2341
	Cheng-Few Lee, Alice C. Lee, and Wen-chi Yeh	
100	Alternative Errors-in-Variables Models and Their Applications in Finance Research	2369
	Hong-Yi Chen, Alice C. Lee, and Cheng-Few Lee	
101	Optimal Payout Ratio under Uncertainty and the Flexibility Hypothesis: Theory and Empirical Evidence	2409
	Cheng-Few Lee, Manak C. Gupta, Hong-Yi Chen, and Alice C. Lee	
102	Mergers and Acquisitions: Principles and Practices	2457
	Harvey A. Poniachek	
103	Accrual Accounting and Risk: Abnormal Sales Growth, Accruals Quality, and Returns	2501
	Liu Min Shirley	
104	Applications of Book-Tax Difference in Accounting and Finance Research	2539
	Nan-Ting Kuo and Shang-En Yu	
105	Evaluating Portfolio Risk Management: A New Evidence from DCC Models and Wavelet Approach	2557
	Rabeh Khalfaoui, Aviral Kumar Tiwari, and Xuan Vinh VO	
106	Cash Conversion Cycle and Corporate Performance: Global Evidence	2597
	Chong-Chuo Chang	
107	How Consistent Are the Judges of Portfolio Performance?	2627
	Matthew Brigida, Chin W. Yang, and Ken Hung	
108	Risk Aversion and the Value of Information for Investors	2633
	Richard E. Kihlstrom	
109	A Fuzzy Real Option Valuation Approach To Capital Budgeting Under Uncertainty Environment	2655
	Shin-Yun Wang and Cheng-Few Lee	

Part III Appendices	2679
Appendix A: Derivation of Dividend Discount Model	2681
Appendix B: Derivation of DOL, DFL and DCL	2683
Appendix C: Derivation of Crossover Rate	2685
Appendix D: Capital Budgeting Decisions with Different Lives	2687
Appendix E: Derivation of Minimum-Variance Portfolio	2691
Appendix F: Derivation of an Optimal Weight Portfolio Using the Sharpe Performance Measure	2693
Appendix G: Applications of the Binomial Distribution to Evaluate Call Options	2697
Appendix H: Derivation of Modigliani and Miller (M&M) Proposition I and II with Taxes	2707
Appendix I: Derivation of Capital Market Line (CML)	2711
Appendix J: Derivation of Capital Market Line (SML)	2713
Appendix K: Derivation of Black-Scholes Option Pricing Model	2715
Appendix L: Hillier's Statistical Distribution Method	2719
Appendix M: Capital-Rationing Decision	2727
Appendix N: Noncentral χ^2 and the Option Pricing Model	2731
Appendix O: Impacts of Financing Decisions on Capital Budgeting Decisions	2735
Index	2741

About the Editors

Professor Cheng-Few Lee is a Distinguished Professor of Finance at Rutgers Business School, Rutgers University, and was chairperson of the Department of Finance from 1988 to 1995. He has also served on the faculty of the University of Illinois (IBE Professor of Finance) and the University of Georgia. He has maintained academic and consulting ties in Taiwan, Hong Kong, China, and the United States for the past three decades. He has been a consultant to many prominent groups including, the American Insurance Group, the World Bank, the United Nations, The Marmon Group Inc., Wintek Corporation, and Polaris Financial Group.

Professor Lee founded the *Review of Quantitative Finance and Accounting* (RQFA) in 1990 and the *Review of Pacific Basin Financial Markets and Policies* (RPBFMP) in 1998, and serves as managing editor for both journals. He was also a co-editor of the *Financial Review* (1985–1991) and the *Quarterly Review of Economics and Finance* (1987–1989). In the past 47 years, Dr. Lee has written numerous textbooks ranging in subject matters from financial management to corporate finance, security analysis and portfolio management to financial analysis, planning and forecasting, and business statistics. In addition, he edited two popular books, *Encyclopedia of Finance* (with Alice C. Lee) and *Handbook of Quantitative Finance and Risk Management* (with Alice C. Lee and John Lee). Dr. Lee has also published more than 220 articles in more than 20 different journals in finance, accounting, economics, statistics, and management. Professor Lee was ranked the most published finance professor worldwide during the period 1953–2008.

Professor Lee was the intellectual force behind the creation of the new Masters of Quantitative Finance program at Rutgers University. This program began in 2001 and has been ranked as one of the top 15 quantitative finance programs in the United States. Professor Lee started the Conference on Financial Economics and Accounting. This conference is a consortium of Rutgers University, New York University, Temple University, University of Maryland, Georgia State University, Tulane University, Indiana University, and University of Toronto. This conference is the most well-known conference in finance and accounting.

John C. Lee is a Microsoft Certified Professional in Microsoft Visual Basic and Microsoft Excel VBA. He holds a bachelor's and master's degree in accounting from the University of Illinois at Urbana-Champaign.

John has worked for over 20 years in both the business and technical fields as an accountant, auditor, systems analyst, and business software developer. He is the author of the book on how to use MINITAB and Microsoft Excel for statistical analysis, which is a companion text to *Statistics of Business and Financial Economics*, 2nd and 3rd, of which he is one of the coauthors. In addition, he has also coauthored the textbooks *Financial Analysis, Planning and Forecasting*, 2nd ed. (with Cheng F. Lee and Alice C. Lee), and *Security Analysis, Portfolio Management, and Financial Derivatives* (with Cheng F. Lee, Joseph Finnerty, Alice C. Lee, and Donald Wort). John has been a Senior Technology Officer at the Chase Manhattan Bank and Assistant Vice President at Merrill Lynch. Currently, he is the Director of the Center for PBBEF Research.

Contributors

Zvika Afik Management of Service Organizations Department, Hadassah Academic College, Jerusalem, Israel

James S. Ang Department of Finance, College of Business, Florida State University, Tallahassee, FL, USA

James R. Barth Auburn University and Milken Institute, Auburn, AL, USA
Department of Finance, Harbert College of Business, Auburn University, Auburn, AL, USA

John Blin APT Investment Management, Hong Kong, China

Matthew Brigida SUNY Polytechnic Institute, Utica, NY, USA

John Cadle University of Birmingham, Birmingham, UK

Rachel Calipha School of Management and Economics, The Academic College of Tel Aviv-Yaffo, Tel Aviv, Israel

Chong-Chuo Chang Department of Banking and Finance, National Chi Nan University, Puli Township, Taiwan

Hao Chang Jacobs Levy Equity Management, Florham Park, NJ, USA

Jow-Ran Chang National Tsing Hua University, Hsinchu, Taiwan, Republic of China

Hong-Yi Chen Department of Finance, National Chengchi University, Taipei, Taiwan

Ren-Raw Chen Fordham University, New York, NY, USA

Sheng-Syan Chen Department of Finance, College of Commerce, National Chengchi University, Taipei City, Taiwan

Wei Chen School of Business, Fuzhou Institute of Technology, Fuzhou, China

Yingmei Cheng Department of Finance, Florida State University, Tallahassee, FL, USA

Thomas C. Chiang Department of Finance, Drexel University, Philadelphia, PA, USA

Michael T. Chng School of Accounting, Economics and Finance, Deakin University, Melbourne, VIC, Australia

Kyung-Jin Choi Korea University Business School, Seoul, South Korea

Quentin C. Chu University of Memphis, Memphis, TN, USA

Michael Dothan Atkinson Graduate School of Management, Willamette University, Salem, OR, USA

Chang-Wen Duan Tamkang University, New Taipei, Taiwan, Republic of China

Wayne E. Ferson University of Southern California, Los Angeles, CA, USA

Joseph E. Finnerty University of Illinois, Champaign, IL, USA

Jonathan Fletcher Department of Accounting and Finance, University of Strathclyde, Glasgow, UK

Iraj Fooladi Rowe School of Business, Dalhousie University, Halifax, NS, Canada

John A. Fox Matthew Brigida, Clarion University of Pennsylvania, Clarion, PA, USA

Reto Francioni Swiss Stock Exchange, Zurich, Switzerland

Jack Clark Francis Economics and Finance, Baruch College, CUNY, New York, NY, USA

Gerard L. Gannon School of Accounting, Economics and Finance, Deakin University, Melbourne, VIC, Australia

Feng Gao Department of Accounting and Information Systems, Rutgers University, New Brunswick, NJ, USA

Abdul Ghafoor Department of Finance, School of Business, Monash University Malaysia, Bandar Sunway, Selangor, Malaysia

Fernando Gómez-Bezares Deusto Business School, Bilbao, Spain

Fernando R. Gómez-Bezares Mutua Madrileña, Madrid, Spain

Aron Gottesman Department of Finance and Graduate Economics, Lubin School of Business, Pace University, New York, NY, USA

John Guerard McKinley Capital Management, LLC, Computational Finance and Risk Management Program, The University of Washington, Anchorage, AK, USA

Xu Guo College of Economics, Shenzhen University, Shenzhen, China

Manak C. Gupta Temple University, Philadelphia, PA, USA

Wolfgang Karl Härdle Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

Singapore Management University, Singapore Department of Mathematics and Physics, Singapore, Singapore

Charles University in Prague, Prague, Czech Republic

National Yang Ming Chiao Tung University (NYCU), Hsinchu City, Taiwan

Tanweer Hasan Independent University of Bangladesh, Dhaka, Bangladesh

Guanming He Department of Accounting, Durham University, Durham, UK

Yan He Indiana University Southeast, New Albany, IN, USA

Randall A. Heron Indiana University, Indianapolis, IN, USA

Lan-chih Ho Central Bank of the Republic of China (Taiwan), Taipei, Taiwan, People's Republic of China

Thomas S. Y. Ho Thomas Ho Company, Ltd., New York, NY, USA

Cheng Hsiao University of Southern California, Los Angeles, CA, USA

Chang-Tseh Hsieh University of Southern Mississippi, Hattiesburg, MS, USA

Wenhsin Hsu Department of Accounting, National Taiwan University, Taipei, Taiwan

Hanxin Hu Department of Accounting and Information Systems, Rutgers University, Newark, NJ, USA

Jing-Zhi Huang Penn State University, Philadelphia, PA, USA

Ken Hung Division of International Banking and Finance Studies, Texas A&M International University, Laredo, TX, USA

Mao-Wei Hung National Taiwan University, Taipei, Taiwan, Republic of China

Michael B. Imerman Princeton University, Princeton, NJ, USA

Gady Jacoby I.H. Asper School of Business, The University of Manitoba, Winnipeg, MB, Canada

Bikki Jaggi School of Business, Rutgers University, Newark-New Brunswick, NJ, USA

Gwoduan David Jou CFO, Taikang Life Insurance Co., Beijing, China

Woo-Young Kang Department of Economics and Finance, Brunel University London, Uxbridge, UK

Lie-Jane Kao Department of Accounting Information, Takming University of Technology, Taipei, Taiwan

Yiannis Karavias Department of Economics, University of Birmingham, Birmingham, UK

Khondkar E. Karim Rochester Institute of Technology, Rochester, NY, USA

Rabeh Khalfaoui Laboratoire de recherche en Économie et Gestion: LR18ES27, Faculté des Sciences Economiques et de Gestion de Sfax, University of Sfax, Sfax, Tunisia

Richard E. Kihlstrom The Wharton School of the University of Pennsylvania, Philadelphia, PA, USA

Dongcheol Kim Korea University Business School & SolBridge International School of Business at Woosong University, Seoul, South Korea

Soon-Ho Kim Pukyong National University, Busan, South Korea

Richard J. Kish Lehigh University, Bethlehem, PA, USA

Hsien-Chang Kuo National Chi-Nan University, Nantou, Taiwan, Republic of China

Nan-Ting Kuo Shandong University of Technology, Business School, Zibo, Shandong, China

Wikil Kwak School of Accounting, University of Nebraska at Omaha, Omaha, NE, USA

Tsong-Yue Lai Department of Finance, California State University, Fullerton, Fullerton, CA, USA

Alice C. Lee Center for PBBEF Research, Phoenix, AZ, USA

Cheng-Few Lee Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

George H. Lee Beyondbond, Inc, New York, NY, USA

Kang Bok Lee Department of Systems and Technology, Harbert College of Business, Auburn University, Auburn, AL, USA

Sang Bin Lee Hanyang University, Seoul, South Korea

Haim Levy Hebrew University, Jerusalem, Israel

Wilbur G. Lewellen Purdue University, Indianapolis, IN, USA

April Zhichao Li Department of Accounting, Durham University, Durham, UK

Hui Li Illinois Institute of Technology, Chicago, IL, USA

Fang-Pang Lin National Research Laboratories, National Center for High Performance Computing, Hsinchu, Taiwan

Fu-Lai Lin Department of Finance, Dar-Yeh University, Changhua, Taiwan

James Juichia Lin Department of Information Management and Finance, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

Lin Lin National Chi-Nan University, Nantou, Taiwan, Republic of China

William T. Lin Tamkang University, New Taipei, Taiwan, Republic of China

Melody Lo University of Texas at San Antonio, San Antonio, TX, USA

John M. Longo Rutgers Business School – Newark & New Brunswick, Piscataway, NJ, USA

Chiuling Lu Department of International Business, National Taiwan University, Taipei, Taiwan

Andrew Mark GlobeFlex, LP, San Diego, CA, USA

Ryan McDonough Department of Accounting and Information Systems, Rutgers University, Piscataway, NJ, USA

Dwight B. Means Clarion University of Pennsylvania, Clarion, PA, USA

L. Melody University of Texas at San Antonio, San Antonio, TX, USA

Liu Min Shirley Department of Accounting, University of Illinois, Chicago, IL, USA

Shashidhar Murthy Indian Institute of Management Bangalore, Bangalore, India

Nguyen Nguyen Minnesota State University, Mankato, MN, USA

Joseph P. Ogden State University of New York at Buffalo, Buffalo, NY, USA

Oded Palmon Rutgers Business, Rutgers University, Piscataway, NJ, USA
Rutgers Business, Rutgers University, Newark and New Brunswick, NJ, USA

Robert H. Patrick Professor Emeritus, Finance and Economics, Rutgers University, Newark, NJ, USA

Lin Peng Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

Fai-Nan Perng Central Bank of the Republic of China (Taiwan), Taipei, Taiwan, Republic of China

Jenifer Piesse University of London, London, UK

Deborah N. Pittman Rhodes College, Memphis, Tennessee, USA

Anna Pomeranets Bank of Canada, Ottawa, ON, Canada

Harvey A. Poniachek Department of Finance and Economics, Rutgers University, Newark, NJ, USA

Gordon Roberts Schulich School of Business, York University, Toronto, ON, Canada

Nadezhda Safronova School of Economics and Business Engineering, Institute of Econometrics, Statistics and Mathematical Finance, University of Karlsruhe, Karlsruhe, Germany

Savita A. Sahay Department of Accounting and Information Systems, Rutgers Business School, Rutgers University, Piscataway, NJ, USA

Lalith P. Samarakoon University of St. Thomas, St. Paul, MN, USA

Robert A. Schwartz Baruch College of the City University of New York, New York, NY, USA

Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

G. V. Satya Sekhar Gandhi Institute of Technology and Management Studies, GITAM, University, Visakhapatnam, AP, India

Xuan Shen Quantitative Risk Analytics, Regions Bank, Birmingham, AL, USA

Dongxiao Shen Department of Accounting, Durham University, Durham, UK

Yong Shi Information Systems and Quantitative Analysis, University of Nebraska at Omaha, Omaha, NE, USA

Chinese Academy of Sciences, Beijing, China

Keshab Shrestha Department of Finance, School of Business, Monash University Malaysia, Bandar Sunway, Selangor, Malaysia

Jacky Yuk-Chow So Department of Accounting and Finance, Macau University of Science and Technology, Taipa, Macau

Bruno Spilak IRTG 1792, School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

Suresh Srivastava University of Alaska Anchorage, Anchorage, AK, USA

Ting Sun Department of Accounting and Information Systems, The College of New Jersey, Ewing, NJ, USA

Yuhsein Tai Department of Finance and Economics, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

Pearl Tan Singapore Management University, Singapore, Singapore

Nicholas S. P. Tay University of San Francisco, San Francisco, CA, USA

C. H. Ted Hong Beyondbond, Inc, New York, NY, USA

Michael Theobald University of Birmingham, Birmingham, UK

Aviral Kumar Tiwari Rajagiri Business School, Rajagiri Valley Campus, Kochi, India

South Ural State University, Chelyabinsk, Russian Federation

Institute of Business Research, University of Economics Ho Chi Minh City, Ho Chi Minh City, Vietnam

Phuong Dung Tong Ministry of Finance, Corporate Finance, Hanoi, Vietnam

Khairy Tourk Illinois Institute of Technology, Chicago, IL, USA

Tony Sio-Chong U Department of Accounting and Finance, Macau University of Science and Technology, Taipa, Macau

Geraldo M. Vasconcellos Lehigh University, Bethlehem, PA, USA

Itzhak Venezia School of Business, The Hebrew University, Jerusalem, Israel

The Hebrew University of Jerusalem, Jerusalem, Israel

The Academic College of Tel Aviv-Yaffo, Tel Aviv, Israel

Xuan Vinh VO Institute of Business Research, University of Economics Ho Chi Minh City, Ho Chi Minh City, Vietnam

Shin-Huei Wang University of Southern California, Los Angeles, CA, USA

The Center for Operations Research and Econometrics (CORE), Louvain-la-Neuve, Belgium

Shin-Yun Wang Department of Finance, National Dong Hwa University, Hualien, Taiwan

Wen-Ching Wang Robeco Investment Management, New York, NY, USA

Daniel G. Weaver Rutgers University, New Brunswick, NJ, USA

K. C. John Wei School of Accounting and Finance, The Hong Kong Polytechnic University, Kowloon, Hong Kong

James Winder Rutgers University, New Brunswick, NJ, USA

Chou-Yen Wu Department of Accounting, Feng Chia University, Taichung, Taiwan

Chunchi Wu School of Management, University at Buffalo, Buffalo, NY, USA

Po-Cheng Wu Department of Finance and Banking, Kainan University, Taoyuan, Taiwan

Ying Wu Department of Economics and Finance, Franklin P. Perdue School of Business, Salisbury University, Salisbury, MD, USA

Jiayi Xu Mount St. Joseph University, Cincinnati, OH, USA

Yan Xue Rutgers University, Piscataway, NJ, USA

Chin W. Yang Clarion University of Pennsylvania, Clarion, PA, USA
National Chung Cheng University, Chia-yi, Taiwan

Wen-chi Yeh Rutgers University, New Brunswick, NJ, USA

Gillian Yeo Nanyang Technological University, Singapore, Singapore

Yeo Song Yoon Department of Finance, Harbert College of Business, Auburn University, Auburn, AL, USA

Hai-Chin Yu Department of International Business, Chung Yuan University, Chung-Li, Taiwan

Shang-En Yu Finance and Economics College, Jimei University, Jimei, China

Linti Zhang School of Accounting and Finance, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Shen Zhang Troy University, Troy, AL, USA

Zhaohui Zhang Long Island University, Brookville, NY, USA

Part I

Terms and Essays

Cheng-Few Lee and Alice C. Lee

Abstract

In general, topics of finance cover financial accounting, financial management, corporate finance, financial planning and forecasting, futures, options, risk management, international finance, real estate, financial econometrics, financial mathematics, financial statistics, and others. In this portion of the encyclopedia, we present important terms and essays on all these topics in detail. We believe this portion will be useful for both academicians and partitioner.

Keywords

Earnings before interest and taxes (EBIT) · Accounts receivable · Cash conversion cycle · London Interbank Offered Rate (LIBOR) · Capital budgeting projects

1.1 A**1. Abnormal Return**

Return on a stock beyond what would be the expected return that is predicted by market movements alone. [See also **Cumulative abnormal return (CAR)**]

C.-F. Lee (✉)

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: cflee@business.rutgers.edu

A. C. Lee

Center for PBBEF Research, Phoenix, AZ, USA

2. Absolute Cost Advantage

Absolute cost advantages can place competitors at a cost disadvantage, even if the scale of operations is similar for both firms. Such cost advantages can arise from an advanced position along the learning curve, where average costs decline as cumulative output rises over time. This differs from economies of scale, which involves the relationship between average costs and the output level per period of time. A firm that enters a market segment early can learn about the production and distribution process first and make more efficient use of assets, technology, raw inputs, and personnel than its competitors. In such cases, the firm can frequently reduce costs and prices and maintain market leadership. Similar advantages can result from possessing proprietary technology that is protected by patents.

Some firms seek to maintain absolute cost advantages by entering foreign market. Early entry can allow the firm to gain experience over its competitors, as it can more efficiently track foreign market trends and technologies and disseminate new methods throughout the firm.

3. Absolute Priority of Claims

In cases of liquidation of a firm's assets, the rule requires satisfaction of certain claims prior to the satisfaction of other claims. The priority of claims in liquidation or reorganization typically takes the following order:

1. Special current debt, which includes trustee expenses, unpaid wages that employees have earned in the 90 days preceding bankruptcy (not to exceed \$2,000 for any one case), and contributions to employee benefit plans that have fallen due within the 180 days preceding bankruptcy.
2. Consumer claims on deposits not exceeding \$900 per claim.
3. Tax claims.
4. Secured creditors' claims, such as mortgage bonds and collateral trust bonds, but only to the extent of the liquidating value of the pledged assets.
5. General creditors' claims, including amounts owed to unsatisfied secured creditors and all unsecured creditors, but only to the extent of their proportionate interests in the aggregate claims of their classes.
6. Preferred stockholders' claims, to the extent provided in their contracts, plus unpaid dividends.
7. Residual claims of common stockholders.

The priority of claims order and amounts are arbitrary and no conclusions should be drawn about the relative merits of how workers, consumers, the government, creditors, and owners are treated.

4. Absolute Priority Rule (APR)

Establishes priority of claims under liquidation. Once the corporation is determined to be bankrupt, liquidation takes place. The distribution of the proceeds of the

liquidation occurs according to the following priority: (1) administration expenses; (2) unsecured claims arising after the filing of an involuntary bankruptcy petition; (3) wages, salaries, and commissions; (4) contributions to employee benefit plans arising within 180 days before the filing date; (5) consumer claims; (6) tax claims; (7) secured and unsecured creditors' claims; (8) preferred stockholders' claims; (9) common stockholders' claims. APR is similar to absolute priority of claims.

5. Absolute Purchasing Power Parity

Absolute purchasing power parity states that exchange rates should adjust to keep purchasing power constant across currencies. In general, however, absolute purchasing power parity does not hold, in part because of transportation costs, tariffs, quotas, and other free trade restrictions. A more useful offshoot of absolute purchasing power parity is **relative purchasing power parity**. [See also **Relative purchasing power parity**]

6. Accelerated Cost Recovery System (ACRS)

A system used to depreciate accelerated assets for tax purposes. The current system, enacted by the 1986 Tax Reform Act, is very similar to ACRS established in 1981. The current modified cost recovery system specifies the depreciable lives (recovery periods) and rates for each of several classes of property. It should be noted that this higher level of depreciation is offset by reclassifying individual assets into categories with longer life. [See **Modified cost recovery system**]

7. Accelerated Depreciation

A method of computing depreciation deductions for income taxes that permits deductions in early years greater than those under straight line depreciation. It includes sums of year's digits, units of production, and double decline methods. [See these three methods discussed later in sections S, U, and D]

8. Account Activity

Transactions associated with a deposit account, including home debits, transit checks, deposits, and account maintenance.

9. Account Analysis

An analytical procedure for determining whether a customer's deposit account or entire credit-deposit relationship with a bank is profitable. The procedure compares revenues from the account with the cost of providing services.

10. Account Executive

A representative of a brokerage firm who processes orders to buy and sell stocks, options, etc. for a customer's account.

11. Account Maintenance

The overhead cost associated with collecting information and mailing periodic statements to depositors.

12. Accounting, Relationship to Finance

The accounting function, quantifies, to a certain extent, the economic relationships within the firm and provides data on which management bases its planning, controlling, and operating decisions. Like accounting, finance deals with value and the monetary resources of the organization. [See **Finance**]

13. Accounting-Based Beta Forecasting

Elgers (1980, pp. 389–408) proposed accounting-based beta forecasting. Accounting-based beta forecasts rely upon the relationship of accounting information such as the growth rate of the firm, earnings before interest and tax (EBIT), leverage, and the dividend pay-out as a basis for forecasting beta. To use accounting information in beta forecasting, the historical beta estimates are first cross-sectionally related to accounting information such as growth rate, variance of EBIT, leverage, and accounting beta:

$$\beta_i = a_0 + a_1X_{1i} + a_2X_{2i} + a_jX_{ji} + \cdots a_mX_{mi}$$

where β_i is the beta coefficient for i th firm which is estimated in terms of market model. X_{ji} is the j th accounting variables for i th firm, and a_j is the regression coefficient.

14. Accounting-Based Performance Measures

To evaluate firm performance we can use accounting-based measure such as sales, earnings per share, growth rate of a firm. However, accounting performance measures are vulnerable to distortion by accounting principles, whose application may be somewhat subjective (such as when to recognize revenue or how quickly to depreciate assets). Rather than present an unbiased view of firm performance, accounting statements may be oriented toward the perspective that management wants to present. Additionally, accounting-based performance measures are always *historical*, telling us where the firm has been.

15. Accounting Analytic

The use of financial ratios and fundamental analysis to estimate firm specific credit quality examining items such as leverage and coverage measures, with an evaluation of the level and stability of earnings and cash flows. [See **Credit scoring model**]

16. Accounting Beta

Project betas can be estimated based on accounting beta. Accounting measures of return, such as EBIT/Total Assets, can be regressed against a profitability index that is based on data for the stocks in the S&P 500 or some other market index:

$$\left[\frac{EBIT}{TA}\right]_{\text{project},i,t} = \alpha_i + A\beta_i \left[\frac{EBIT}{TA}\right]_{\text{market},t} + \varepsilon_{i,t}$$

where the slope estimate, $A\beta_i$ is the accounting beta.

Accounting information by product line or division is available in various Securities and Exchange Commission (SEC) filings that are required of publicly traded firms. Although a firm's multidivisional structure may disqualify it from being a pure play comparable, it may include divisional data in its public SEC filing that would be useful for estimating an accounting beta.

17. Accounting Break-Even

Accounting break-even occurs when accounting revenues equal accounting expenses so that pretax income (and hence net income) equals zero. It tells us how much product must be sold so that the firm's overall accounting profits are equal to accounting expenses. Ignoring working capital effects,

$$\text{OCF} = \text{NI} + \text{Depreciation.}$$

At accounting break-even, net income (NI) is zero, so operating cash flow (OCF) equals the periodic depreciation expense. Substituting this into the general break-even (Q^*) formula, we obtain accounting break-even quantity ($Q^*_{\text{accounting}}$) as:

$$Q^*_{\text{accounting}} = \frac{FC + Dep}{p - vc},$$

where:

FC = fixed cost

vc = variable cost per unit

p = price per unit

Dep = depreciation

The denominator, $(p - vc)$, is called the contribution margin. The accounting break-even quantity is given by the sum of the fixed cost and depreciation divided by the contribution margin. Accounting break-even tells us how much product must be sold so that the firm's overall accounting profits are not reduced.

18. Accounting Earnings

Earnings of a firm as reported on its income statement. Accounting earnings are affected by several conventions regarding the valuation of assets such as inventories (e.g., LIFO versus FIFO treatment) and by the way some expenditures such as capital investments are recognized over time (such as depreciation expenses).

19. Accounting Income

Income described in terms of accounting earnings, based upon records of transactions in company books kept according to generally accepted principles (GAAP). Accountants generally measure revenues and expenses based on accruals and deferrals rather than cash flows and in turn, measure the net income of the firm by matching its revenues with the costs it incurred to generate those revenues.

Theoretically, financial analysis should consider **economic income** rather than accounting earnings to determine the value of the firm, since economic income represents the firm's true earnings and cash flows. [See also **Economic income**] However, since economic income is not directly observable, analysts generally use accounting earnings as a proxy. The relationship between economic income and accounting earnings can be related by the following equation:

$$\begin{aligned} \text{Accounting Income} &= \text{Economic Income} \\ &(\text{permanent component}) + \text{Error}(\text{Transitory component}) \end{aligned}$$

20. Accounting Insolvency

Total book liabilities exceed total book value of assets. A firm with negative net worth is insolvent on the books.

21. Accounting Liquidity

The ease and quickness with which assets can be converted to cash. Current assets are the most liquid and include cash and those assets that will be turned into cash within a year from the date of the balance sheet. Fixed assets are the least liquid kind of assets.

22. Accounts Payable

Money the firm owes to suppliers. These are payments for goods or services, such as raw materials. These payments will generally be made after purchases. Purchases will depend on the sales forecast. Accounts payable is an unfunded short-term debt.

23. Accounting Rate of Return (ARR)

The accounting rate of return (ARR) method (which is one of the methods for capital budgeting decision) computes a rate of return for a project based on a ratio of average project income to investment outlay (usually either the total initial investment or the average investment is used). Projects with accounting returns exceeding a management-determined minimum return are accepted; those with returns below the cutoff are rejected. To compute the accounting rate of return, we use the following ratio:

$$ARR = \frac{\text{Average annual net income}}{\text{Total initial investment}}$$

Similar to the **payback method**, the accounting rate of return method has none of the four desired selection method characteristics. [See also **Payback method**] First, it does not even use cash flows; it relies on accounting income. Second, it ignores time value of money concepts. Third, it states no clearly defined, objective decision criterion; like the payback method, its cutoff depends on the discretion of management. Fourth, ARR tells us absolutely nothing about the impact of a project on shareholder wealth.

24. Accounts Receivable

Money owed to the firm by customers; the amounts not yet collected from customers for goods or services sold to them (after adjustment for potential bad debts).

25. Accounts Receivable Financing

A secured short-term loan that involves either the assigning of receivables or the factoring of receivables. Under assignment, the lender has a lien on the receivables and recourse to the borrower. Factoring involves the sale of accounts receivable. Then the purchaser, call the factor, must collect on receivables. [See **Factoring**]

26. Accounts Receivable Turnover

Credit sales divided by average accounts receivable. In general, a higher accounts receivable turnover ratio suggests more frequent payment of receivables by customers. The accounts receivable turnover ratio is written as:

$$\text{Accounts receivable turnover} = \frac{\text{Sales}}{\text{Accounts receivable}}$$

Thus, if a firm's accounts receivable turnover ratio is larger than the industry average, this implies that the firm's accounts receivable are more efficiently managed than the average firm in that industry.

27. Accreting Swap

A swap where the notional amount increases over the life of the swap. It is used to hedge interest rate risk or agreements with a rising principal value, such as a construction loan.

28. Accrual

The accumulation of income earned or expense incurred, regardless of when the underlying cash flow is actually received or paid.

29. Accrual Bond

A bond that accrues interest but does not pay interest to the investor until maturity when accrued interest is paid with the principal outstanding.

30. Accrual Earnings Management

Accrual earnings management refers to companies that manipulate accrual-related accounts, such as accounts receivable to increase their earnings number. Accrual-based earnings management aims to obscure true economic performance by changing accounting methods or estimates within the generally accepted accounting principles. Real earnings management alters the execution of real business transactions.

31. Accrual Swap

An interest rate swap where interest on one side accrues only when the floating reference rate is within certain range. The range can be maintained, fixed, or reset periodically during the entire life of swap.

32. Accrued Interest

Interest income that is earned but not yet received. Alternatively, it refers to pro-rated portion of a bond's coupon payment (c) since the previous coupon date with $(m - d)$ days have been passed since last coupon payment, the accrued interest is $c(m - d)/m$ where m and d represent total days and days left to receive coupon payment respectively. In a semiannual coupon, if $m = 182$ days, $d = 91$ days, and $c = \$60$, then the accrued interest is

$$(\$30) \left(\frac{182 - 91}{182} \right) = \$15$$

33. Accumulated Benefit Obligation (ABO)

FASB Statement 87 specifies that the measure of corporate pension liabilities to be used on the corporate balance sheet in external reports is the accumulated benefit

obligation (ABO), which is the present value of pension benefits owed to employees under the plan's benefit formula absent any salary projections and discounted at a nominal rate of interest.

34. Accumulation Phase

During the accumulation phase, the investor contributes money periodically to one or more open-end mutual funds and accumulates shares. [See also **Variable annuities**]

35. Acid-Test Ratio

A measure of liquidity from reported balance sheet figures with targeted minimum value of one. Calculated as the sum of cash, marketable securities, and accounts receivable divided by current liabilities. [See also **Quick ratio**]

36. Acquisition

Assuming there are two firms, Firm A and Firm B. Acquisition is a form of business combination in which Firm B buys Firm A and they both remain in existence, Firm B as the parent and Firm A as the subsidiary.

Mergers or acquisitions are also ways for a private firm to raise equity capital by selling all or part of the firm to another corporation. [See also **Merger**] Another firm may pay an attractive price for the equity of the private firm, especially if the private firm has a good strategic fit with the buyer's products and plans, or if the purchase offers a foreign corporation easy entry into the US market. Acquisitions can be negotiated to allow the firm's managers to retain their current positions or to receive lucrative consulting contracts.

Another advantage to a merger or acquisition is when the investor is a large corporation with deep pockets and a willingness to help the firm grow. Such a situation can provide financing for the firm's present and foreseeable future needs. Rather than spending time canvassing banks and equity investors for capital, management can concentrate on doing what it presumably does best: managing the firm to make it grow and succeed.

The drawback to a merger or acquisition is a loss of control. Although a seemingly straightforward consequence, this can be a large stumbling block for a business with a tradition of family ownership or for a group of founding entrepreneurs who consider the firm their "baby." Unless the private equity owners get an exceptional deal from the new owner, a merger or sale causes them to give up the return potential of their business. If the company does grow and succeed after the sale, someone else – the new investor – will reap the benefits. If the original owners stay with the new owner, they may become frustrated by the lack of attention from their new partners if the firm is only a small part of the acquirer's overall business.

37. Active Bond Portfolio Management

An investment policy whereby managers buy and sell securities prior to final maturity to speculate on future interest rate movements. In addition, managers can also identify the relative mispricing within the fixed-income market.

38. Active Management.

Attempts to achieve portfolio returns more than commensurate with risk, either by forecasting broad market trends or by identifying particular mispriced sectors of a market or securities in a market.

39. Active Portfolio

In the context of the Treynor-Black model (see Treynor and Black 1973), the portfolio formed by mixing analyzed stocks of perceived nonzero alpha values. This portfolio is ultimately mixed with the passive market index portfolio. [See **Alpha and active portfolio management**]

40. Activity Charge

A service charge based on the number of checks written by a depositor.

41. Activity Ratios

Activity ratios measure how well a firm is using its resources. Four activity ratios are analyzed: (1) **inventory turnover**, (2) **average collection period**, (3) **fixed-asset turnover**, and (4) **total asset turnover**.

Inventory turnover (sales/inventory) measures how well a firm is turning over its inventory. The average collection period (receivables/sales per day) measures the accounts-receivable turnover. The fixed-asset turnover (sales to net fixed assets) measures the turnover of plant and equipment – a measure of capacity utilization. Total-asset turnover (sales/total assets) measures how efficiently the total asset has been utilized.

42. Acts of Bankruptcy

Bankruptcy includes a range of court procedures in the United States that may result in the firm being liquidated or financially reorganized to continue operations. This may occur voluntarily if the firm permits a petition for bankruptcy, or a creditor's petition may force the firm into the courts. Such a petition by a creditor charges the firm with committing one of the following acts of bankruptcy: (1) committing fraud while legally insolvent, (2) making preferential disposition of firm assets while legally insolvent, (3) assigning assets to a third party for voluntary liquidation while insolvent, (4) failing to remove a lien on the firm within 30 days

while insolvent, (5) appointment of a receiver or trustee while insolvent, or (6) written admission of insolvency.

43. Actual Maturity

The number of days, months, or years between today and the date a loan or security is redeemed or retired.

44. Add-On Interest

Add-on interest means that the total interest owed on the loan, based on the annual stated interest rate, is added to the initial principal balance before determining the periodic payment. This kind of loan is called add-on loan. Payments are determined by dividing the total of the principal plus interest by the number of payments to be made. When a borrower repays a loan in a single, lump sum, this method gives a rate identical to annual stated interest. However, when two or more payments are to be made, this method results in an effective rate of interest that is greater than the nominal rate. Putting this into equation form, we see that:

$$PV = \sum_{t=1}^N \frac{\text{Future flows}}{(1 + \text{Interest rate})^t}$$

where:

PV = the present value or loan amount

t = the time period when the interest and principal repayment occur

N = the number of periods

For example, if the million-dollar loan were repaid in two 6-month installments of \$575,000 each, the effective rate would be higher than 15%, since the borrower does not have the use of the funds for the entire year. Allowing r to equal the annual percentage rate of the loan, we obtain the following:

$$\$1,000,000 = \frac{\$575,000}{(1 + \frac{r}{2})^1} + \frac{\$575,000}{(1 + \frac{r}{2})^2}$$

Using a financial calculator, we see that r equals 19.692% which is also annual percentage return (APR). Using this information we can obtain the installment loan amortization schedule as presented in Table 1.

45. Add-On Rate

A method of calculating interest charges by applying the quoted rate to the entire amount advanced to a borrower times the number of financing periods. For example, an 8% add-on rate indicates \$80 interest per \$1,000 for 1 year, \$160 for 2 years, and

Table 1 Installment loan amortization schedule

	(a)	(b)	(c)	(d)	(e)
			Interest	Principal paid	Ending loan balance
Period	Payment	Beginning balance	$(0.19692)/2 \times (b)$	$(a) - (c)$	$(b) - (d)$
1	\$575,000	\$1,000,000	\$98.460	\$476,540	\$523,460
2	575,000	523,460	51,540	523,460	0
Biannual payment		\$575,000			
Initial balance		\$1,000,000			
Initial maturity		1 year			
APR		19.692%			

so forth. The effective interest rate is higher than the add-on rate because the borrower makes installment payments and cannot use the entire loan proceeds for full maturity. [See **Add-on interest**]

46. Additions to Net Working Capital

Component of cash flow of firm, along with operating cash flow and capital spending. Cash flows that used for making investments in net working capital.

$$\begin{aligned} \text{Total cash flow of the firm} &= \text{Operating cash flow} \\ &\quad - \text{Capital spending} - \text{Additions to net working capital} \end{aligned}$$

47. Additive vs. Multiplicative Models

The additive model is useful when the seasonal variation is relatively constant over time, while the multiplicative model is useful when the seasonal variation increases over time. Additive relationships mean you add the same number to any x-value to get the corresponding y-value. If it is not the same number every time, it is not an additive relationship. Subtraction can be an additive relationship because subtracting a number is the same as adding a negative number. Multiplicative relationships mean you multiply any x-value times the same number to get the corresponding y-value. If it is not the same number every time, it is not multiplicative. Division can be a multiplicative relationship because dividing by a number is the same as multiplying by its reciprocal (flip the number).

48. Adjustable Mortgage Instrument (AMI)

A home mortgage loan under which some of the terms of the loan, such as the contract loan rate or the maturity (term) of the loan, will vary as financial market conditions change.

49. Adjustable-Rate Mortgage (ARM)

A mortgage whose interest rate varies according to some specified measure of the current market interest rate. The adjustable-rate contract shifts much of the risk of fluctuations in interest rates from the lender to the borrower.

50. Adjusted Beta

The sample beta estimated by market model can be modified by using cross-sectional market information (see Vasicek Oldrich 1973, pp. 1233–1239). This kind of modified beta is called adjusted beta. Merrill Lynch's adjusted beta is defined as

$$\text{Adjusted beta} = \frac{2}{3} \text{sample beta} + \frac{1}{3}(1)$$

51. Adjusted Forecast

A (micro or macro) forecast that has been adjusted for the imprecision of the forecast. When we forecast GDP or interest rate over time, we need to adjust for the imprecision of the forecast of either GDP or interest rate.

52. Adjusted Price Value (APV) Model

Adjusted present value model for capital budgeting decision. This is one of the methods used to do capital budgeting for levered firm. This method takes into account the tax shield value associated with tax deduction for interest expense. The formula can be written as

$$APV = NPV + T_c D$$

where:

APV = Adjusted present value

NPV = Net present value

T_c = Marginal corporate tax rate

D = Total corporate debt

$T_c D$ = Tax shield value

This method is based upon M&M Proposition I tax. [See **M&M proposition I with tax**]

53. ADR

American Depositary Receipt: A certificate issued by a US bank which evidences ownership in foreign shares of stock held by the bank. [See **American depository receipt**]

54. Advance

A payment to a borrower under a loan agreement.

55. Advance Commitment

This is one of the methods for hedging interest rate risk in a real estate transaction. It is a promise to sell an asset before the seller has lined up purchase of the asset. This seller can offset risk by purchasing a futures contract to fix the sale price. We call this a long hedge by a mortgage banker because the mortgage banker offsets risk in the cash market by buying a future contract.

56. Affiliate

Any organization that is owned or controlled by a bank or bank holding company, the stockholders, or executive officers.

57. Affinity Card

A credit card that is offered to all individuals who are part of a common group or who share a common bond.

58. Aftermarket

The period of time following the initial sale of securities to the public, this may last from several days to several months.

59. After-Acquired Clause

A first mortgage indenture may include an after-acquired clause. Such a provision states that any property purchased after the bond issue is considered to be security for the bondholders' claim against the firm. Such a clause also often states that only a certain percentage of the new property can be debt financed.

60. After-Tax Real Return

The after-tax rate of return on an asset minus the rate of inflation.

61. After-Tax Salvage Value

After-tax salvage value can be defined as:

$$\text{After-tax salvage value} = \text{Price} - T(\text{Price} - BV),$$

where $Price$ = market value; T = corporate tax rate; and BV = book value.

If $T(Price - BV)$ is positive, the firm owes taxes, reducing the after-tax proceeds of the asset sale; if $T(Price - BV)$ is negative, the firm reduces its tax bill, in essence increasing the after-tax proceeds of the sale. When $T(Price - BV)$ is zero, no tax adjustment is necessary.

By their nature, after-tax salvage values are difficult to estimate as both the salvage value and the expected future tax rate are uncertain.

As a practical matter, if the project termination is many years in the future, the present value of the salvage proceeds will be small and inconsequential to the analysis. If necessary, however, analysts can try to develop salvage value forecasts in two ways. First, they can tap the expertise of those involved in secondary market uses of the asset. Second, they can try to forecast future scrap material prices for the asset. Typically, the after-tax salvage value cash flow is calculated using the firm's current tax rate as an estimate for the future tax rate.

The problem of estimating values in the distant future becomes worse when the project involves a major strategic investment that the firm expects to maintain over a long period of time. In such a situation, the firm may estimate annual cash flows for a number of years and then attempt to estimate the project's value as a going concern at the end of this time horizon. One method the firm can use to estimate the project's going-concern value is the **constant dividend growth model**. [See also **Constant dividend growth model** or **Gordon model**]

62. After-Tax WACC Method

The after-tax, weighted-average, cost-of-capital method has two noticeable differences from the formulation of the equity-residual method. First, no flows associated with the debt financing appear in the numerator, or as relevant cash flows. Second, the cost of capital is adjusted downward, with k_w being a weighted average of debt and equity costs, the debt expense accounted for on an after-tax basis. In that way debt financing is reckoned with in an indirect manner. With the assumption that r and k_e are constant over time, k_w can be affected only by the debt-to-equity ratio, a problem most often avoided by assuming a fixed debt-to-equity ratio. See Appendix O for further detail.

63. Agency Bond

Bonds issued by federal agencies such as Government National Mortgage Association (GNMA) and government-sponsored enterprises such as Small Business Administration (SBA). Agency bond is a direct obligation of the Treasury even though some agencies are government sponsored or guaranteed. The net effect is that agency bonds are considered almost default-risk free (if not legally so in all cases) and, therefore, are typically priced to provide only a slightly higher yield than their corresponding T-bond counterparts.

64. Agency Costs

The principal-agent problem imposes agency costs on shareholders. Agency costs are the tangible and intangible expenses borne by shareholders because of the self-serving actions of managers. Agency costs can be explicit, out-of-pocket expenses (sometimes called **direct agency costs**) or more implicit ones (sometimes called **implicit agency costs**). [See also **Principal-agent problem**]

Examples of explicit agency costs include the costs of auditing financial statements to verify their accuracy, the purchase of liability insurance for board members and top managers, and the monitoring of managers' actions by the board or by independent consultants.

Implicit agency costs include restrictions placed against managerial actions (e.g., the requirement of shareholder votes for some major decisions) and covenants or restrictions placed on the firm by a lender.

The end result of self-serving behaviors by management and shareholder attempts to limit them is a reduction in firm value. Investors will not pay as much for the firm's stock because they realize that the principal-agent problem and its attendant costs lower the firm's value. Agency costs will decline, and firm value will rise, as principals' trust and confidence in their agents rises.

Alternately, costs of conflicts of interest among stock holders, bondholders, and managers. Agency costs are the costs of resolving these conflicts. They include the costs of providing managers with an incentive to maximize shareholder wealth and then monitoring their behavior, and the cost of protecting bondholders from shareholders. Agency costs are borne by stockholders.

65. Agency Costs, Across National Borders

Agency costs may differ across national borders as a result of different accounting principles, banking structures, and securities laws and regulations. Firms in the United States and the United Kingdom use relatively more equity financing than firms in France, Germany, and Japan. Some argue that these apparent differences can be explained by differences in equity and debt agency costs across the countries.

For example, agency costs of equity seem to be lower in the United States and the United Kingdom. These countries have more accurate systems of accounting (in that the income statements and balance sheets are higher quality reflecting actual revenues and expenses, assets and liabilities) than the other countries, and have higher auditing standards. Dividends and financial statements are distributed to shareholders more frequently, as well, which allows shareholders to monitor management more easily. Germany, France, and Japan, on the other hand, all have systems of debt finance that may reduce the agency costs of lending. In other countries, a bank can hold an equity stake in a corporation, meet the bulk of the corporation's borrowing needs, and have representation on the corporate board of directors. Corporations can own stock in other companies and also have representatives on other companies' boards. Companies frequently get financial advice from groups of banks and other large corporations with whom they have interlocking directorates. These

institutional arrangements greatly reduce the monitoring and agency costs of debt; thus, debt ratios are substantially higher in France, Germany, and Japan.

66. Agency Problem

Conflicts of interest among stockholders, bondholders, and managers.

67. Agency Securities

Fixed-income securities issued by agencies owned or sponsored by the federal government. The most common securities are issued by the Federal Home Loan Bank, Federal National Mortgage Association, and Farm Credit System.

68. Agency Theory

The theory of the relationship between principals and agents. It involves the nature of the costs of resolving conflicts of interest between principals and agents. [See also **Agency cost**]

69. Agents

Agents are representatives of insurers. There are two systems used to distribute-or sell-insurance. The **direct writer system** involves an agent representing a single insurer, whereas the **independent agent system** involves an agent representing multiple insurers. An independent agent is responsible for running an agency and for the operating costs associated with it. Independent agents are compensated through commissions, but direct writers may receive either commissions or salaries.

70. Aggregation

This is a process in long-term financial planning. It refers to the smaller investment proposals of each of the firm's operational units are added up and in effect treated as a big picture.

71. Aging Accounts Receivable

A procedure for analyzing a firm's accounts receivable by dividing them into groups according to whether they are current or 30, 60, or over 90 days past due. [See **Aging schedule**]

72. Aging Population

A long-term increase in the average age of people in many countries is resulting in changes in their saving habits and in their demands for financial services, thus putting added pressure on financial institutions to develop new financial services.

73. Aging Schedule of Account Receivable

A compilation of accounts receivable by the age of account.

Typically, this relationship is evaluated by using the average collection period ratio. This type of analysis can be extended by constructing an aging-of-accounts-receivable table, such as Table 2. The table shows an example of decline in the quality of accounts receivable from January to February as relatively more accounts have been outstanding for 61 days or longer. This breakdown allows analysis of the cross-sectional composition of accounts over time. A deeper analysis can assess the risk associated with specific accounts receivable, broken down by customer to associate the probability of payment with the dollar amount owed.

74. Allocational Efficiency

The overall concept of allocational efficiency is one in which security prices are set in such a way that investment capital is directed to its optimal use. Because of the position of the United States in the world economy, the international and domestic efficiency, also, since the overall concept of allocational efficiency is too general to test, operational efficiency must be focused upon as a testable concept.

75. Allowance for Loan and Lease Losses

An accounting reserve set aside to equate expected (mean) losses from credit defaults. It is common to consider this reserve as the buffer for expected losses and some risk-based economic capital as the buffer for unexpected losses.

76. All-in-Cost

The weighted average cost of funds for a bank calculated by making adjustments for required reserves and deposit insurance costs, the sum of explicit and implicit costs.

Table 2 Aging schedule of accounting receivable

Days outstanding	January		February	
	Accounts receivable range	Percent of total	Accounts receivable range	Percent of total
0–30 days	\$250,000	25.0%	\$250,000	22.7%
31–60 days	500,000	50.0	525,000	47.7
61–90 days	200,000	20.0	250,000	22.7
Over 90 days	50,000	5.0	75,000	6.8
Total accounts receivable	\$1,000,000	100.0%	\$1,100,100	100.0%

77. Alpha

The abnormal rate of return on a security in excess of what would be predicted by an equilibrium model like CAPM or APT. For CAPM, the alpha for the i th firm (α_i) can be defined as:

$$\alpha_i = (\bar{R}_i - R_f) - \beta_i(\bar{R}_m - R_f)$$

where:

$\bar{R}_i$ = average return for the i th security

$\bar{R}_m$ = average market rate of return

R_f = risk-free rate

β_i = beta coefficient for the i th security

Treynor and Black (1973) have used the alpha value to form active portfolio.

78. Alternative Minimum Tax (AMT)

A federal tax against income intended to ensure that taxpayers pay some tax even when they use tax shelters to shield income.

79. American Depositary Receipt (ADR)

A security issued in the United States to represent shares of a foreign stock, enabling that stock to be traded in the United States. For example, Taiwan Semiconductor (TSM) from Taiwan has sold ADR in the United States.

80. American Option

An American option is an option that can be exercised at any time up to the expiration date. The factors that determine the values of American and **European options** are the same except the time to exercise the option; all other things being equal, however, an American option is worth more than a European option because of the extra flexibility it grants the option holder. [See also **European option**]

81. Amortize

To reduce a debt gradually by making equal periodic payments that cover interest and principal owed. In other words, it liquidates on an installment basis. [See also **Amortization**]

82. Amortization

Repayment of a loan in installments. Long-term debt is typically repaid in regular amounts over the life of the debt. At the end of the amortization the entire

indebtedness is said to be extinguished. Amortization is typically arranged by a **sinking fund**. Each year the corporation places money into a sinking fund, and the money is used to buy back the bond.

83. Amortization Schedule for a Fixed-Rate Mortgage

Amortization schedule for a fixed-rate mortgage is used to calculate either the monthly or the annual payment for a fixed rate mortgage. The following example is used to show the procedure for calculating annual payment for a fixed-rate mortgage:

Suppose Bill and Debbie have taken out a home equity loan of \$5,000 which they plan to repay over 3 years. The interest rate charged by the bank is 10%. For simplicity, assume that Bill and Debbie will make annual payments on their loan. (a) Determine the annual payments necessary to repay the loan. (b) Construct a loan amortization schedule.

- (a) Finding the annual payment requires the use of the present value of an annuity relationship:

$$\begin{aligned} PVAN &= (\$CF) \left[\frac{1 - \left(\frac{1}{1+r} \right)^n}{r} \right] = (\$CF) \left[\frac{1 - \left(\frac{1}{1+.10} \right)^3}{.10} \right] \\ &= \$5000 = (\$CF)(2.48685) \end{aligned}$$

This result is an annual payment of $\$5,000/2.48685 = \$2,010.57$.

- (b) Below is the loan amortization schedule constructed for Bill and Debbie:

(1)	(2)	(3)	(4)	(5)	(6)
			Interest paid	Principal paid	Ending balance
Year	Beginning balance	Annuity payments	(2) $\times$ 0.10	(3) $-$ (4)	(2) $-$ (5)
1	\$5,000.00	\$2,010.57	\$500.00	\$1,510.57	\$3,489.43
2	3,489.43	2,010.57	348.94	1,661.63	1,827.80
3	1,827.80	2,010.57	182.78	1,827.79	0.01

84. Amortizing Swap

An interest rate swap in which the outstanding notional principal amount declines over time. It generally is used to hedge interest rate risk or mortgage or other amortized loan.

85. Analyst Coverage Network

The number of analysts who observe a particular stock or other security. Large and well-known companies often have greater analyst coverage of their stocks than other companies. Greater analyst coverage often results in higher demand and therefore lower prices. Some value investors recommend buying securities with low analyst coverage because they could be undervalued.

86. Analyst Recommendation Revisions

Financial analysts recommend stock or option investments after using technical or fundamental analysis. When the situation changes, they will revise their recommendations.

87. Analysts' Information

Financial analysts use past and present financial data to help their company determine solid financial plans. They constantly adapt and learn alongside changing market conditions in order to forecast investment opportunities.

88. Angels

Individuals providing venture capital. These investors do not belong to any venture-capital firm; these investors act as individuals when providing financing. However, they should not be viewed as isolated investors.

89. Announcement Date

Date on which particular news concerning a given company is announced to the public. Used in **event studies**, which researchers use to evaluate the economic impact of events of interest. For example, dividend announcement date. [See also **Event studies**]

90. Announcement Effect

The effect on stock returns for the first trading day following an event announcement. For example, earnings announcement and dividend announcement will affect the stock price.

91. Annual Effective Yield

Also called the **effective annual rate (EAR)**. [See also **Effective annual rate (EAR)**]

92. Annual Percentage Rate (APR)

Banks, finance companies, and other lenders are required by law to disclose their borrowing interest rates to their customers. Such a rate is called a **contract** or **stated rate**, or more frequently, an annual percentage rate (APR). The method of calculating the APR on a loan is preset by law. The APR is the interest rate charged per period multiplied by the number of periods in a year:

$$\text{APR} = r \times m,$$

where:

r = periodic interest charge

m = number of periods per year

However, the APR misstates the true interest rate. Since interest compounds, the APR formula will understate the true or effective interest cost. The **effective annual rate (EAR)**, sometimes called the *annual effective yield*, adjusts the APR to take into account the effects of compounded interest over time. [See also **Effective annual rate (EAR)**]

It is useful to distinguish between a contractual or stated interest rate and the group of rates we call yields, effective rates, or market rate. A contract rate, such as the annual percentage rate (APR), is an expression that is used to specify interest cash flows such as those in loans, mortgages, or bank savings accounts. The yield or effective rate, such as the effective annual rate (EAR), measures the opportunity costs; it is the true measure of the return or cost of a financial instrument.

93. Annualized Holding-Period Return

The annual rate of return that when compounded T times would have given the same T -period holding return as actual occurred from period 1 to period T . If R_t is the return in year t (express in decimals), then $(1 + R_1) \times (1 + R_2) \times (1 + R_3) \times (1 + R_4)$ is called a 4-year holding period return.

94. Annuity

An annuity is a series of consecutive, equal cash flows over time. In a regular annuity, the cash flows are assumed to occur at the *end* of each time period. Examples of financial situations that involve equal cash flows include fixed interest payments on a bond, and cash flows that may arise from insurance contracts, retirement plans, and amortized loans such as car loans and home mortgages.

The future value of an n -period annuity of \$C per period is:

$$FVAN = \$C \left[1 + (1 + r) + (1 + r)^2 + (1 + r)^3 + \dots + (1 + r)^{n-1} \right]$$

which can be reduced to:

$$FVAN = \$C \left[\frac{(1 + r)^n - 1}{r} \right] = \$C \times FVIFA(r, n),$$

where $FVIFA(r, n)$ represents the future value interest factor for an annuity.

To find the present value of an n -period annuity of $\$CF$ per period is:

$$PVAN = \$C \left[\frac{1}{(1 + r)} + \frac{1}{(1 + r)^2} + \frac{1}{(1 + r)^3} + \dots + \frac{1}{(1 + r)^n} \right]$$

which can be shown as:

$$PVAN = \$C \left[\frac{1}{r} - \frac{1}{(1 + r)^n} \right] = \$CF \times PVIFA(r, n),$$

where $PVIFA(r, n)$ is the present value interest factor for an annuity.

95. Annuity Due

When a cash flow occurs at the *beginning* of each annuity period, the annuity becomes an annuity due. Since the cash flows in the n -year annuity due occurs at the beginning of each year, they are invested for one extra period of time compared to the n -year regular annuity. This means all the annuity due cash flows are invested at r percent interest for an extra year.

To take this one extra year of compounding into account, the future value interest factor for an annuity [$FVIFA(r, n)$] can be multiplied by $(1 + r)$ to determine the future value interest factor for an annuity due (FVANDUE):

$$\begin{aligned} FVANDUE &= \$C \left[\frac{(1 + r)^n - 1}{r} \right] (1 + r) \\ &= \$C * FVIFA(r, n) * (1 + r) \end{aligned}$$

Many situations also require present value calculations for cash flows that occur at the beginning of each time period. Examples include retirement checks that arrive on the first of the month and insurance premiums that are due on the first of the month. Again, the cash flows for the n -year annuity due occur 1 year *earlier* than those of the n -year regular annuity, making them more valuable. As in determining the FVANDUE, we can adjust for this simply by multiplying the corresponding PVIFA by $(1 + r)$ to reflect the fact that the cash flows are received one period sooner in an annuity due. The formula for the present value of an annuity due (PVANDUE) is:

$$\begin{aligned} \text{PVANDUE} &= \$C \left[\frac{1 - \left(\frac{1}{1+r}\right)^n}{r} \right] * (1+r) \\ &= \$C * \text{PVIFA}(r, n) * (1+r) \end{aligned}$$

96. Annuity Factor

The term used to calculate the present value or future value of the stream of level payments for a fixed period. [See **Annuity** for detail]

97. Annuity in Advance

An annuity with an immediate initial payment. This is called annuity due. [See also **Annuity due**]

98. Annuity in Arrears

An annuity with a first payment one full period, rather than immediately. That is, the first payment occurs on date 1 rather than on date 0.

99. Anticipated Income Theory

A theory that the timing of loan payments should be tied to the timing of a borrower's expected income.

100. Antithetic Variate Method

A technique used in Monte Carlo valuation, in which each random draw is used to create two simulated prices from opposite tails of the asset price distribution. This is one of the variance reduction procedures. Other method is stratified sampling method. [See **Stratified sampling**]

101. Applied Research

A research and development (R&D) component that is riskier than **development projects**. [See also **Development projects**] It seeks to add to the firm's knowledge base by applying new knowledge to commercial purposes.

102. Appraisal Ratio

The signal-to-noise ratio of an analyst's forecasts. The ratio of alpha to residual standard deviation. This ratio measures abnormal return per unit of risk that in principle could be diversified away by holding a market index portfolio.

103. Appraisal Rights

Rights of shareholders of an acquired firm that allow them to demand that their shares be purchased at a fair value by the acquiring firm.

104. Appreciation

An increase in the market value of an asset. For example, you buy one share of IBM stock at \$90. After 1 year you saw the stock for \$100, then this investment appreciated by 11.11%.

105. Appropriation Phase of Capital Budgeting

The focus of the appropriation phase, sometimes called the *development* or *selection phase*, is to appraise the projects uncovered during the identification phase. After examining numerous firm and economic factors, the firm will develop estimates of expected cash flows for each project under examination. Once cash flows have been estimated, the firm can apply time value of money techniques to determine which projects will increase shareholder wealth the most.

The appropriation phase begins with information generation, which is probably the most difficult and costly part of the phase. Information generation develops three types of data: internal financial data, external economic and political data, and nonfinancial data. This data supports forecasts of firm-specific financial data, which are then used to estimate a project's cash flows. Depending upon the size and scope of the project, a variety of data items may need to be gathered in the information generation stage. Many economic influences can directly impact the success of a project by affecting sales revenues, costs, exchange rates, and overall project cash flows. Regulatory trends and political environment factors, both in the domestic and foreign economies, also may help or hinder the success of proposed projects.

Financial data relevant to the project is developed from sources such as marketing research, production analysis, and economic analysis. Using the firm's research resources and internal data, analysts estimate the cost of the investment, working capital needs, projected cash flows, and financing costs. If public information is available on competitors' lines of business, this also needs to be incorporated into the analysis to help estimate potential cash flows and to determine the effects of the project on the competition.

Nonfinancial information relevant to the cash flow estimation process includes data on the various means that may be used to distribute products to consumers, the quality and quantity of the domestic or nondomestic labor forces, the dynamics of technological change in the targeted market, and information from a strategic analysis of competitors. Analysts should assess the strengths and weaknesses of competitors and how they will react if the firm undertakes its own project.

After identifying potentially wealth-enhancing projects, a written proposal, sometimes called a *request for appropriation* is developed and submitted to the manager with the authority to approve. In general, a typical request for appropriation requires an executive summary of the proposal, a detailed analysis of the project, and data to support the analysis.

The meat of the appropriation request lies in the detailed analysis. It usually includes sections dealing with the need for the project, the problem or opportunity that the project addresses, how the project fits with top management's stated objectives and goals for the firm, and any impact the project may have on other operations of the firm.

The appropriation process concludes with a decision. Based upon the analysis, top management decides which projects appear most likely to enhance shareholder wealth. The decision criterion should incorporate the firm's primary goal of maximizing shareholder wealth.

106. Arbitrage

Arbitrage is when traders buy and sell virtually identical assets in two different markets in order to profit from price differences between those markets.

Besides currencies, traders watch for price differences and arbitrage opportunities in a number of financial markets, including stock markets and futures and options markets. In the real world, this process is complicated by trading commissions, taxes on profits, and government restrictions on currency transfers. The vigorous activity in the foreign exchange markets and the number of traders actively seeking risk-free profits prevents arbitrage opportunities based on cross-rate mispricing from persisting for long.

In other words arbitrage refers to buying an asset in one market at a lower price and simultaneously selling an identical asset in another market at a higher price. This is done with no cost or risk.

107. Arbitrage Condition

Suppose there are two riskless assets offering rates of return r and r' , respectively. Assuming no transaction costs, one of the strongest statements that can be made in positive economic is that

$$r = r' \quad (1)$$

This is based on the law of one price, which says that the same good cannot sell at different prices. In terms of securities, the law of one price says that securities which identical risks must have the same expected return. Essentially, Eq. 1 is an arbitrage condition that must be expected to hold in all but the most extreme circumstances. This is because if $r > r'$, the first riskless asset could be purchased with funds obtained from selling the second riskless asset. This arbitrage transaction would yield a return of $r - r'$ without having to make any new investment of funds or take

on any additional risk. In the process of buying the first asset and selling the second, investors would bid up the former's price and bid down the latter's price. This repricing mechanism would continue up to the point where these two assets' respective prices equaled each other. And thus $r = r'$

108. Arbitrage Pricing Theory (APT)

S. Ross (1976) derived a generalized capital asset pricing relationship called the arbitrage pricing theory (APT). To derive the APT, Ross assumed the expected rate of return on asset i at time t , $E(R_{it})$, could be explained by k independent influences (or factors):

$$E(R_{it}) = \alpha + \beta_{i1}(\text{factor } 1) + \beta_{i2}(\text{factor } 2) + \dots + \beta_{ik}(\text{factor } k),$$

where β_{ik} measures the sensitivity of the i th asset's returns to changes in factor k (sometimes called index k). In the terminology of factor analysis, β_{ik} 's are called factor loading

Using the prior equation, Ross shows that the actual return of the i th security can be defined as:

$$R_i = E(R_i) + [F_1 - E(F_1)]\beta_{i1} + \dots + [F_k - E(F_k)]\beta_{ik},$$

where $[F_k - E(F_k)]$ represents the surprise or change in the k th factor brought about by systematic economic events.

Like the **capital asset pricing model (CAPM)**, the APT assumes that investors hold diversified portfolios, so only systematic risks affect returns. [See also **Capital asset pricing model (CAPM)**] The APT's major difference from the CAPM is that it allows for more than one systematic risk factor. The APT is a *generalized* capital asset pricing model; the CAPM is a special, one-factor case of the APT, where the one factor is specified to be the return on the market portfolio.

The APT does have a major practical drawback. It gives no information about the specific factors that drive returns. In fact, the APT does not even tell us how many factors there are. Thus, testing the APT is purely empirical, with little theory to guide researchers. Estimates of the number of factors range from two to six; some studies conclude that the market portfolio return is one of the return-generating factors, while others do not. Some studies conclude that the CAPM does a better job in estimating returns; others conclude that APT is superior.

The jury is still out on the superiority of the APT over the CAPM. Even though the APT is a very intuitive and elegant theory and requires much less restrictive assumptions than the CAPM, it currently has little practical use. It is both difficult to determine the return-generating factors and to test the theory.

In sum, an equilibrium asset pricing theory that is derived from a factor model by using diversification and arbitrage. It shows that the expected return on any risky asset is a linear combination of various factors.

109. Arbitrageur

An individual engaging in arbitrage. [See also **Arbitrage**]

110. Arithmetic Average

The risk of an item is reflected in its variability from its average level. For comparison, a stock analyst may want to determine the level of return and the variability in returns for a number of assets to see whether investors in the higher risk assets earned a higher return over time. A financial analyst may want to examine historical differences between risk and profit on different types of new product introductions or projects undertaken in different countries.

If historical, or ex-post, data are known, an analyst can easily compute historical average return and risk measures. If X_t represent a data item for period t , the arithmetic average $\bar{X}$, over n periods is given by:

$$\bar{X} = \frac{\sum_{t=1}^n X_t}{n}$$

In sum, the sum of the values observed divided by the total number of observation – sometimes referred to as the mean. [See also **Geometric average**]

111. ARCH Model

In econometrics, the autoregressive conditional heteroskedasticity (ARCH) model is a statistical model for time series data that describes the variance of the current error term or innovation as a function of the actual sizes of the previous time periods' error terms; often the variance is related to the squares of the previous innovations. This model has been frequently used in finance to estimate the Johnson hedge ratio.

112. Arditti-Levy WACC method

The Arditti-Levy method is most similar to the after-tax weighted-average cost-of-capital method. It was seen as necessary to restate the after-tax, weighted-average, cost-of-capital formula in this manner because the tax payment to the government has an influence on the net cash flows, and for that reason the cash-flow figures would be misleading. To rectify this problem the discount rate must now be adjusted since the interest tax shield is reflected in the cash flows and double counting would be involved. While the after-tax, weighted-average cost of capital recognized the cost of debt in the discount rate, it was akin to the equity-residual method in considering only returns to equity. The Arditti and Levy formulas imply that a weighted average discount rate including the lower cost of debt could only (or best) be used if all flows to all sources of financing were included; hence the term rDt is found at the end of the first term. This can be rationalized if one considers

the case of a firm where there is one owner. The total cash flow to the owner is the relevant figure, and the discount rate applicable is simply a weighted average of the two individual required rates of return. See Appendix O for further detail.

113. ARIMA Model

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). In finance, the ARIMA model is frequently used to forecast EPS or stock price per share.

114. Arithmetic Mean

[See **Arithmetic average**]

115. ARM

Adjustable rate mortgage is a mortgage in which the contractual interest rate is tied to some index of interest rates (prime rate for example) and charges when supply and demand conditions change the underlying index. [See also **Adjustable rate mortgage**]

116. Arrears

An overdue outstanding debt. In addition, we use arrearage to indicate the overdue payment.

117. Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) or connectionist systems are computing systems inspired by the biological neural networks that constitute animal brains. Such systems learn (progressively improve their ability) to do tasks by considering examples, generally without task-specific programming. ANNs are frequently used to forecast the chance of credit card default.

118. Asian Option

An option in which the payoff at maturity depends upon an average of the asset prices over the life of the option.

119. Asian Tail

A reference price that is computed as an average of recent prices. For example, an equity-linked note may have a payoff based on the average daily stock price over the last 20 days (the Asian tail).

120. Ask Price

The price at which a dealer or market-maker offers to sell a security. Also called the *offer price*.

121. Asked Price

The price at which a securities dealer is willing to sell securities held in his or her portfolio to the public.

122. Assets

Anything that the firm owns. It includes current, fixed, and other assets. Asset can also be classified as tangible and intangible assets.

123. Asset Allocation Decision

Choosing among broad asset classes such as stocks versus bonds. In other words, asset allocation is an approach to investing that focuses on determining the mixture of asset classes that is most likely to provide a combination to risk and expected return that is optimal for the investor. In addition to this, portfolio insurance is an asset-allocation or hedging strategy that allows the investor to alter the amount of risk he or she is willing to accept by giving up some return.

124. Asset-Backed Debt Securities (ABS)

Issuers of credit have begun following the lead set by mortgage lenders by using asset securitization as a means of raising funds. Securitization meaning that the firm repackages its assets and sells them to the market.

In general, an ABS comes through certificates issued by a grantor trust, which also registers the security issue under the Securities Act of 1933. These securities are sold to investors through underwritten public offerings or private placements. Each certificate represents a fractional interest in one or more pools of assets. The selling firm transfers assets, with or without recourse, to the grantor trust, which is formed and owned by the investors, in exchange for the proceeds from the certificates. The trustee receives the operating cash flows from the assets and pays scheduled interest and principal payments to investors, servicing fees to the selling firm, and other expenses of the trust.

From a legal perspective, the trust owns the assets that underlie such securities. These assets will not be consolidated into the estate of the selling firm if it enters into bankruptcy.

To date, most ABS issues have securitized automobile and credit-card receivables. It is expected that this area will grow into other fields, such as computer leases, truck leases, land and property leases, mortgages on plant and equipment, and commercial loans.

125. Asset-Backed Security

A security with promised principal and interest payments backed or collateralized by cash flows originated from a portfolio of assets that generate the cash flows.

126. Asset-Based Financing

Financing in which the lender relies primarily on cash flows generated by the asset financed to repay the loan.

127. Asset-Liability Management

The management of a bank's entire balance sheet to achieve desired risk-return objectives and to maximize the market value of stockholders' equity. Asset-liability management is the management of the net interest margin to ensure that its level and riskiness are compatible with risk/return objectives of the institution.

128. Asset Management Ratios

Asset management ratios (also called activity or *asset utilization ratios*) attempt to measure the efficiency with which a firm uses its assets.

Receivables Ratios

Accounts receivable turnover ratio is computed as credit sales divided by accounts receivable. In general, a higher accounts receivable turnover ratio suggests more frequent payment of receivables by customers. The accounts receivable turnover ratio is written as:

$$\text{Accounts receivable turnover} = \frac{\text{Sales}}{\text{Accounts receivable}}$$

Thus, if a firm's accounts receivable turnover ratio is larger than the industry average; this implies that the firm's accounts receivable are more efficiently managed than the average firm in that industry.

Dividing annual sales by 365 days gives a daily sales figure. Dividing accounts receivable by daily sales gives another asset management ratio, the **average collection period of credit sales**. In general, financial managers prefer shorter collection periods over longer periods.

Comparing the average collection period to the firm's credit terms indicates whether customers are generally paying their accounts on time. The **average collection period** is given by:

$$\text{Average collection period} = \frac{\text{Accounts receivable}}{\text{Sales}/365}$$

The average collection period (ACP) is easy to calculate and can provide valuable information when compared to current credit terms or past trends.

One major drawback to the ACP calculation, however, is its sensitivity to changing patterns of sales. The calculated ACP rises with increases in sales and falls with decreases in sales. Thus, changes in the ACP may give a deceptive picture of a firm's actual payment history. Firms with seasonal sales should be especially careful in analyzing accounts receivable patterns based on ACP. For instance, a constant ACP could hide a longer payment period if it coincides with a decrease in sales volume. In this case, the ACP calculation would fail to properly signal a deterioration in the collection of payments.

Inventory Ratios

The **inventory turnover ratio** is a measure of how quickly the firm sells its inventory. It is computed as cost of goods sold divided by inventory. The ratio clearly depends upon the firm's inventory accounting method: for example, last-in, first-out (LIFO) or first-in, first-out (FIFO). The inventory turnover ratio is written as:

$$\text{Inventory turnover} = \frac{\text{Cost of goods sold}}{\text{Inventory}}$$

It is an easy mistake to assume that higher inventory turnover is a favorable sign; it also may signal danger. An increasing inventory turnover may raise the possibility of costly stockouts. Empty shelves can lead to dissatisfied customers and lost sales.

Fixed and Total Assets Ratio

The **total asset turnover ratio** is computed as sales divided by total assets. The **fixed asset turnover ratio** is sales divided by fixed assets. Similar to the other turnover ratio, these ratios indicate the amount of sales generated by a dollar of total and fixed assets, respectively. Although managers generally favor higher fixed and total asset turnover ratios, these ratios can be *too* high. The fixed asset turnover ratio may be large as a result of the firm's use of old, depreciated equipment. This would indicate that the firm's reliance on old technology could hurt its future market position, or that it could face a large, imminent expense for new equipment, including the downtime required to install it and train workers.

A large total asset turnover ratio also can result from the use of old equipment. Or, it might indicate inadequate receivables arising from an overly strict credit system or dangerously low inventories.

The asset turnover ratios are computed as follows:

$$\text{Total asset turnover} = \frac{\text{Sales}}{\text{Total assets}}$$

$$\text{Fixed asset turnover} = \frac{\text{Sales}}{\text{Fixed assets}}$$

129. Asset Pricing

Asset pricing refers to determining the value of bonds, stocks, or options. The asset pricing model includes the capital market line (CML), security market line (SML), and the Black-Scholes option pricing model. In financial economics, asset pricing refers to a formal treatment and development of two main pricing principles, together with the resultant models. There have been many models developed for different situations, but correspondingly, these stem from general equilibrium asset pricing or rational asset pricing, the latter corresponding to risk neutral pricing.

130. Asset Pricing Tests

The asset pricing tests specifically test the hypothesis that pricing errors of the chosen set of test assets is small, in a suitable metric, either in absolute terms or relative to a reference asset (e.g., the risk-free or the “market” portfolio).

131. Assets Requirements

A common element of a financial plan that describes projected capital spending and the proposed uses of net working capital. Asset requirements increase when sales increase.

132. Asset Sensitive

A bank is classified as asset sensitive if its GAP is positive. Under this case interest rate sensitive asset is larger than interest rate sensitive liability.

133. Asset Swap

Effectively transforms an asset into an asset of another type, such as converting a fixed rate bond into a floating-rate bond. Results in what is known as a “synthetic security.”

134. Asset Turnover (ATO)

The annual sales generated by each dollar of assets (sales/assets). It can also be called as asset utilization ratio.

135. Asset-or-Nothing Call

An option that pays a unit of the asset if the asset price exceeds the strike price at expiration or zero otherwise.

136. Asset-or-Nothing Option

An option that pays a unit of the asset if the option is in-the-money or zero otherwise.

137. Assignment

The transfer of the legal right or interest on an asset to another party.

138. Assumable Mortgage

The mortgage contract is transferred from the seller to the seller to the buyer of house.

139. Asymmetric Butterfly Spread

A butterfly spread in which the distance between strike prices is not equal. [See **Butterfly spread**]

140. Asymmetric Information

Asymmetric information, also known as “information failure,” occurs when one party to an economic transaction possesses greater material knowledge than the other party. In general, managers have more information about his or her company relative to investors.

141. As-You-Like-It Option

[See **Chooser option**]

142. At the Money

The owner of a put or call is not obligated to carry out the specified transaction but has the *option* of doing so. If the transaction is carried out, it is said to have been *exercised*. *At the money* means that the stock price is trading at the exercise price of the option.

143. Auction

A method used to sell securities in which buyers file bids and the highest-price bidders receive securities.

144. Auction Market

A market where all traders in a certain good meet at one place to buy or sell and asset. The NYSE is an example for stock auction market.

145. Auction Method

The principal means by which US Treasury securities are sold to the public today.

146. Audit, or Control, Phase of Capital Budgeting Process

The audit, or *control*, phase is the final step of the capital budgeting process for approved projects. In this phase, the analyst tracks the magnitude and timing of expenditures while the project is progressing. A major portion of this phase is the post-audit of the project, through which past decisions are evaluated for the benefit of future project analyses.

Many firms review spending during the control phase of approved projects. Quarterly reports often are required in which the manager overseeing the project summarizes spending to date, compares it to budgeted amounts, and explains differences between the two. Such oversight during this implementation stage slows top managers to foresee cost overruns. Some firms require projects that are expected to exceed their budgets by a certain dollar amount or percentage to file new appropriation requests to secure the additional funds. Implementation audits allow managers to learn about potential trouble areas so future proposals can account for them in their initial analysis. Implementation audits generally also provide top management with information on which managers generally provide the most accurate estimates of project costs.

In addition to implementation costs, firms also should compare forecasted cash flows to actual performance after the project has been completed. This analysis provides data regarding the accuracy over time of cash flow forecasts, which will permit the firm to discover what went right with the project, what went wrong, and why. Audits force management to discover and justify any major deviations of actual performance from forecasted performance. Specific reasons for deviations from the budget are needed for the experience to be helpful to all involved. Such a system also helps to control intra-firm agency problems by helping to reduce “padding” (i.e., overestimating the benefits of favorite or convenient project proposals). This increases the incentives for department heads to manage in ways that will help the firm achieve its goals.

Investment decisions are based on estimates of cash flows and relevant costs, while in some firms the post-audit is based on accrued accounting and assigned overhead concepts. The result is that managers make decisions based on cash flow, while they are evaluated by an accounting-based system.

A concept that appears to help correct this evaluation system problem is **economic value added (EVA)**. [See also **Economic value added (EVA)**]

The control or post-audit phase sometimes requires the firm to consider terminating or abandoning an approved project. The possibility of abandoning an investment prior to the end of its estimated useful or economic life expands the options available to management and reduces the risk associated with decisions based on holding an asset to the end of its economic life. This form of contingency planning gives decision makers a second chance when dealing with the economic and political uncertainties of the future.

147. Audits of Project Cash Flow Estimated

Capital budgeting audits can help the firm learn from experience. By comparing actual and estimated cash flows, the firm can try to improve upon areas in which forecasting accuracy is poor.

In a survey conducted in the late 1980s, researchers found that three-fourths of the responding *Fortune 500* firms audited their cash flow estimates. Nearly all of the firms that performed audits compared initial investment outlay estimates with actual costs; all evaluated operating cash flow estimates; and two-thirds audited salvage-value estimates. About two-thirds of the firms that performed audits claimed that actual initial investment outlay estimates usually were within 10% of forecasts. Only 43% of the firms that performed audits could make the same claim with respect to operating cash flows. Over 30% of the firms confessed that operating cash flow estimates differed from actual performance by 16% or more. This helps to illustrate that our cash flow estimates are merely point estimates of a random variable. Because of their uncertainty, they may take on higher or lower values than their estimated value.

To be successful, the cash flow estimation process requires a commitment by the corporation and its top policy-setting managers; this commitment includes the type of management information system the firm uses to support the estimation process. Past experience in estimating cash flows, requiring cash flow estimates for all projects, and maintaining systematic approaches to cash flow estimation appear to help firms achieve success in accurately forecasting cash flows.

148. Autocorrelation [Serial Correlation]

The correlation of a variable with itself over successive time intervals. The correlation coefficient can be defined as:

$$\rho = \frac{\text{cov}(r_t, r_{t-1})}{\sigma_t \sigma_{t-1}}$$

where $\text{cov}(r_t, r_{t-1})$ is the covariance between r_t , r_{t-1} , and σ_t and σ_{t-1} are standard deviations for r_t and r_{t-1} , respectively.

Two useful empirical examples of autocorrelation are:

Interest rate exhibit mean reversion behavior and are often negatively auto correlated (i.e., an up move 1 day will suggest a down move the next). But note that mean reversion does not technically necessitate negative autocorrelation.

Agency credit ratings typically exhibit move persistence behavior and are positively auto correlated during downgrades (i.e., a downgrade will suggest another downgrade soon). But, for completeness, note that upgrades do not better predict future upgrades.

149. Automated Clearing House System (ACH)

An Automated Clearing House (ACH) system is an information transfer network that joins banks or other financial institutions together to facilitate the transfer of cash

balances. An ACH system has a high initial fixed cost to install but requires a very low variable cost to process each transaction. The Federal Reserve operates the nation's primary ACH, which is owned by the member banks of the Federal Reserve System. Most of the nation's banks are member of an ACH.

Instead of transferring information about payments or receipts via paper documents like checks, an ACH transfers the information electronically via a computer.

150. Automated Clearinghouse

A facility that processes interbank debits and credits electronically.

151. Automated Loan Machine

A machine that serves as a computer terminal and allows a customer to apply for a loan and, if approved, automatically deposits proceeds into an account designated by the customer.

152. Automated Teller Machines (ATM)

The globalization of automated teller machines (ATMs) is one of the newer frontiers for expansion for US financial networks. The current system combines a number of worldwide communication switching networks, each one owned by a different bank or group of banks.

A global ATM network works like a computerized constellation of switches. Each separate bank is part of a regional, national, and international financial system.

After the customer inserts a credit card, punches a personal identification number (PIN), and enters a transaction request, the bank's computer determines that the card is not one of its own credit cards and switches the transaction to a national computer system. The national system, in turn, determines that the card is not one of its own, so it switches to an international network, which routes the request to the US Global Switching Center. The center passes the request to a regional computer system in the United States, which evaluates the request and responds through the switching network. The entire time required for this process, from initiation at the ATM until the response is received, is reassured in seconds. The use, acceptance, and growth of systems like this will revolutionize the way international payments are made well into the twenty-first century.

153. Availability Float

It refers to the time required to clear a check through the banking system. This process takes place by using either Fed-check collection service, corresponding banks, or local clearing house.

154. Average Accounting Return (AAR)

The average project earnings after taxes and depreciation divided by the average book value of the investment during its life. [See **Accounting rate of returns**]

155. Average Annual Yield

A method to calculate interest that incorrectly combines simple interest and compound interest concepts on investments of more than 1 year. For example, suppose you invested \$10,000 in a 5-year CD offering 9.5% interest compounded quarterly, you would have \$15,991.10 in the account at the end of 5 years. Dividing your \$5,991.10 total return by five, the average annual return will be 11.98%.

156. Average Collection Period

Average amount of time required to collect an accounting receivable. Also referred to as days sales outstanding. [See also **Asset management ratios** and **Activity ratios**]

157. Average Cost of Capital

A firm's required payout to the bondholders and the stockholders expressed as a percentage of capital contributed to the firm. Average cost of capital is computed by dividing the total required cost of capital by the total amount of contributed capital. Average cost of capital (ACC) formula can be defined as:

$$ACC = \frac{S}{V}r_E + \frac{B}{V}(1 - \tau_c)i$$

where V = total market value of the firm; S = value of stockholder's equity; B = value of debt; r_E = rate of return of stockholder's equity; i = interest rate on debt; and τ_c = corporate tax rate.

Here, r_E is the cost of equity, and $(1 - \tau_c)i$ is the cost of debt. Hence, r_A is a weighted average of these two costs, with respective weights S/V and B/V .

158. Average Daily Sales

Annual sales divided by 365 days.

159. Average Exposure

Credit exposure arising from market-driven instruments will have an ever-changing mark-to-market exposure amount. The average exposure represents the average of several expected exposure value calculated at different forward points over the life of swap starting from the end of the first year. The expected exposures are weighted by the appropriate discount factors for this average calculation.

160. Average Price Call Option

The payoff of average price call option = $\max[0, A(T) - K]$, where $A(T)$ is the arithmetic average of stock price over time and K is the strike price. This implies that

the payoff of this option is either equal to zero or larger than zero. In other words, the amount of payoff is equal to the difference between $A(T)$ and K .

161. Average Price Put Option

The payoff of average price put option = $\max[0, K - A(T)]$, where $A(T)$ is the arithmetic average of stock price per share over time and K is the strike price. This implies that the payoff of this option is either equal to zero or larger than zero. In other words, the amount of payoff is equal to the difference between K and $A(T)$.

162. Average Shortfall

The expected loss given that a loss occurs, or as the expected loss given that losses exceed a given level.

163. Average Strike Option

An option that provides a payoff dependent on the difference between the final asset price and the average asset price. For example, an average strike call = $\max[0, S_T - A(T)]$, where $A(T)$ represents average stock price per share over time and S_T represents stock price per share in period T .

164. Average Tax Rate

The average tax rate is the tax bill of a firm divided by its earnings before income taxes (i.e., pretax income). For individuals, it is their tax bill divided by their taxable income. In either case, it represents the percentage of total taxable income that is paid in taxes.

165. Averaging Down

Averaging down involves buying more of a stock or an option at a lower price than the original purchase to reduce the average cost. It is an investing strategy that involves a stock owner purchasing additional shares of a previously initiated investment after the price has dropped. The result of this second purchase is a decrease in the average price at which the investor purchased the stock. It may be contrasted with averaging up.

1.2 B

1. Back Testing

Testing a value-at-risk or other model using historical data. For example, under the current BIS market risk-based capital requirements, a bank must back test its internal

market model over a minimum of 250 past days if it is used for capital requirement calculations. If the forecast VAR errors on those 250 days are too large (i.e., risk is underestimated on too many days), a system of penalties is imposed by regulators to create incentives for bankers to get their models right.

2. Back-to-Back Transaction

A transaction where a dealer enters into offsetting transactions with different parties, effectively serving as a go-between.

3. Backward Equation

[See **Kolmogorov backward equation**]

4. Backwardation

A forward curve in which the futures prices are falling with time to expiration.

5. Backwardization

The situation in which futures prices in futures contracts that expire farther in the future are below prices of nearby futures contracts.

6. Backwards Induction

A procedure for working from the end of a tree to its beginning in order to value an option.

7. Bad Debts

Loans that are due but are uncollectible.

8. Balanced-Budget Unit

An individual, business firm, or unit of government whose current expenditures equal its current receipts of income and therefore is neither a borrower nor a lender of funds.

9. Balance Inquiry

A request by a depositor or borrower to obtain the current balance in his or her account.

10. Balance Sheet

The balance sheet provides a static description of the firm's financial position at a fixed point in time. It details the firm's assets and liabilities at the end of the fiscal year for an annual report or at the end of a quarter for a quarterly statement.

The balance comes from a basic accounting equality:

$$\text{Total assets} = \text{Total liabilities} + \text{Total equity}$$

This equation implies that a firm's assets must equal the total of its liabilities and owners' equity. Stated more informally, what the firm owns (assets) equals what it owes (liability claims to creditors plus equity claims to shareholders). The balance sheet shows how all assets are financed, either by borrowing (debt) or owners' investment (equity).

The left-hand side of the balance sheet reports company assets. It divides the total into current assets, plant and equipment, and other assets (which may include such intangible assets as patents and goodwill). The balance sheet lists these categories in order of liquidity. Liquidity is the ability to quickly convert an asset to cash without a loss in value. The most liquid assets, cash and short-term investments of excess cash, such as marketable securities, are listed first; less liquid assets follow.

The right-hand side of the balance sheet shows the claims against company assets. Categories for these claims include current liabilities, long-term debt, common stock, and retained earnings. The liability and equity claims are listed in order of increasing maturity. This order also reflects the general priority of the claims of creditors and equity holders against the firm's cash flows.

11. Balanced Funds

The balanced funds offer a complete investment program to their clients, so far as marketable securities are concerned. Their portfolio are presumably structured to include bonds and stocks in a ratio considered appropriate for an average individual investor given the return outlook for each sector and possibly a risk and volatility constraint.

12. Balance-of-Payments (BOP) Accounts

A double-entry bookkeeping system recording a nation's transactions with other nations, including exports, imports, and capital flows.

13. Balloon Loan

A loan that requires small payments that are insufficient to pay off the entire loan so that a large final payment is necessary at termination.

14. Balloon Payment

Large final payment, as when a loan is repaid in installments. For example, (a) most high-quality bond issues establish payments to the sinking fund that are not sufficient to redeem the entire issue. As a consequence, there is the possibility of a large balloon payment at maturity; (b) if a lease has a schedule of payments that is very high at the start of the lease term and thereafter very low then these early balloon payments would be an evidence that the lease was being used to avoid taxes and not for a legitimate business purpose.

15. Bank-Discount Interest

Bank-discount interest commonly is charged for short-term business loans. Generally, the borrower makes no intermediate payments, and the life of the loan usually is 1 year or less. Interest is calculated on the amount of the loan, and the life of the loan usually is 1 year or less. Interest is calculated on the amount of the loan, and the borrower receives the difference between the amount of the loan and the amount of interest. In the example, this gives an interest rate of 15%. The interest (\$150,000) is subtracted from the \$1 million loan amount and the borrower has the use of \$850,000 for 1 year. Dividing the interest payment by the amount of money actually used by the borrower (\$150,000 divided by \$850,000), we find the effective rate is 17.6%.

16. Bank Discount Method

The procedure by which yields on US Treasury bills, commercial paper, and bankers' acceptances are calculated; a 360-day year is assumed and there is no compounding of interest income.

17. Bank Discount Yield

An annualized interest rate assuming simple interest, a 360-day year, and using the face value of the security rather than purchase price to compute return per dollar invested.

18. Bank Drafts

Bank drafts, or bills of exchange, is a basic instrument of foreign trade financing that allow exporters to use their banks as collection agents for foreign accounts. The bank forwards the exporter's invoices to the foreign buyer, either by mail or through a branch or correspondent bank in the buyer's country. When the buyer pays the draft, the exporter's bank converts the proceeds of the collection into the exporter's currency and deposits this money in the exporter's account. Two kinds of bank drafts include **sight drafts** and **time drafts**. [See also **Sight draft** and **Time draft**]

19. Bank Failure

A bank failure is the closing of an insolvent bank by a federal or state regulator. The comptroller of the currency has the power to close national banks; banking commissioners in the respective states close state-chartered banks. Banks close when they are unable to meet their obligations to depositors and others. When a bank fails, the Federal Deposit Insurance Corporation (FDIC) covers the insured portion of a depositor's balance, including money market accounts.

20. Bank Holding Company

Any firm that owns or controls at least one commercial bank.

21. Bank of Japan Financial Network System

As the Japanese banks have become increasingly more important in international financial flows, their transfer systems also have grown in importance. The Bank of Japan Financial Network System (BOJ-NET) is a cash and securities wire transfer system for yen-denominated payments. The cash wire, an online funds transfer system for banks, is the Japanese counterpart of CHIPS. Financial institutions use BOJ-NET to provide net settlement services for the Japanese clearinghouse system that clears bills and checks. BOJ-NET also provides settlement for the Japanese electronic fund transfer (EFT) system called *Zengin*. Institutions also can use BOJ-NET to settle yen payments that arise from cross-border transfers and foreign-exchange transactions.

22. Bank Regulatory Compliance

Banking compliance means complying with regulations, laws, and guidelines, whether internal or external. Its function is to prevent, detect, and address any and all deviations, illegalities, and nonconformities in the company's operations. Therefore, it is a valuable support tool for managers and people responsible for information processing, security, risk management, ethical conduct, and other controls and data reliability (including related to money laundering).

23. Bank Relationships

The rapport you establish with the bank with whom you conduct business transactions, which could help smooth the way when it comes to loan applications or special requests.

24. Banker's Acceptance

The banker's acceptance is a comparatively specialized credit source largely confined to financing foreign trade (its only major use within the United States has

been in financing purchases of raw cotton crops). One of the major difficulties in conducting business overseas is in accessing the creditworthiness of potential customers. This problem is best solved by getting a bank to add its reputation to that of the buyer by accepting, or endorsing, the note payable. The investment attractiveness of banker's acceptances must be stressed because most investors are unfamiliar with this short-term, liquid high-yielding investment.

Banker's acceptances are **time drafts** drawn on and accepted by banks, usually to secure arrangements between unfamiliar firms. [See also **Time draft**] They are frequently used in international trade. After generating a banker's acceptance, a bank typically sells it to an investor at a discount. Maturities range from 30 to 180 days, while denominations vary from \$25,000 to over \$1 million, depending upon the specific transaction the banker's acceptance was originally created to finance. Banker's acceptances are relatively illiquid compared to T-bills and most carry higher yields than CDs because of the heterogeneous characteristics.

The interest rate on acceptances is quite low, usually at or very slightly above the prime rate. Any bank that performs services of this kind for its customers probably will expect to be compensated in other ways, however, especially through the maintenance of good demand deposit balances.

In sum, Banker's acceptance is an agreement by a bank to pay a given sum of money at a future date. These agreements typically arise when a seller sends a bill or draft to a customer. The customer's bank accepts this bill and notes the acceptance on it, which makes it an obligation of the bank.

25. Bankers Bank

A firm that provides correspondent banking services to commercial banks and not to commercial or retail deposit and loan customers.

26. Banking

Banking is defined as the business activity of accepting and safeguarding money owned by other individuals and entities, and then lending out this money in order to conduct economic activities such as making profit or simply covering operating expenses.

27. Banking Structure

The number, relative sizes, and types of banks and bank services offered in a given market or in the industry as a whole.

28. Bankrupt

The situation in which a borrower is unable to pay obligated debts.

29. Bankruptcy Costs

The major drawback of having debt in the capital structure is its legal requirement for timely payment of interest and principal. As the debt-to-equity ratio rises, or as earnings become more volatile, the firm will face higher borrowing costs, driven upward by bond investors requiring higher yields to compensate for additional risk.

A rational marketplace will evaluate the probability and associated costs of bankruptcy for a levered firm. Bankruptcy costs include explicit expenses, such as legal and accounting fees and court costs, along with implicit costs, such as the use of management time and skills in trying to prevent and escape bankruptcy. It also is difficult to market the firm's products and keep good people on staff when the firm is teetering on the brink of bankruptcy.

The market will evaluate the present value of the expected bankruptcy costs and reduce its estimate of the value of the firm accordingly. When bankruptcy costs are included in an analysis of M&M Proposition I with taxes, the value of the firm is given by:

$$V_L = V_U + (T)(D) - PV(\text{Expected bankruptcy costs})$$

This says that the value of the levered firm equals the value of the unlevered firm plus the present value of the interest tax shield, minus the present value of expected bankruptcy costs. Incorporating bankruptcy costs into M&M Proposition I relationship between firm value and debt reduces the debt-to-equity ratio at which the firm's value is maximized to less than 100% debt financing. According to the **static trade-off hypothesis**, increases in debt beyond this optimal level actually reduce firm value, as investors' perceptions of the increased cost of bankruptcy outweigh the tax benefits of additional debt. [See also **Static trade-off hypothesis**]

In sum, debt puts pressure on the firm, because interest and principal payment are obligations. If these obligations are not met, the firm may risk some sort of financial distress. The ultimate distress is bankruptcy, where ownership of the firm's assets is legally transferred from the stockholders to the bondholders. Bankruptcy costs tend to offset the advantage to debt. [Also see **Financial distress costs**]

30. Bankruptcy Reform Act

A federal law originally passed in 1978 and subsequently amended that made it easier for consumers to file bankruptcy petitions and keep substantial personal assets that cannot be sold or repossessed to repay outstanding debts.

31. Barbell

An investment portfolio in which a large fraction of securities mature near-term and another large fraction of securities mature longer-term.

32. Bargain-Purchase-Price Option

Gives lessee the option to purchase the asset at a price below fair market value when the lease expires.

33. Barrier Option

An option that has a payoff depending upon whether, at some point during the life of the option, the price of the underlying asset has moved past a reference price (the barrier). Examples are knock-in and knock-out options. [See **Knock-in and knock-out options**]

34. Base Case

Incremental cash flows are the anticipated changes in cash flow from the base case. [See also **Incremental cash flows**] The firm's base-case projection must assess what the firm's market share and cash flows would be if *no* new projects were implemented; in other words, the after-tax cash flows *without* the project. The firm's planners must recognize that if nothing is done, customers may start buying competitors' products in response to the marketing, new product development, and/or quality efforts of the competition. The base-case estimate should reflect these potential declines in cash flow.

35. Base Rate

An interest rate used as an index to price loans; typically associated with a bank's weighted marginal cost of funds.

36. Basel Committee on Banking Supervision

The Basel Committee on Banking Supervision (BCBS) is a committee of banking supervisory authorities that was established by the central bank governors of the Group of Ten countries in 1974. The committee expanded its membership in 2009 and then again in 2014. In 2019, the BCBS has 45 members from 28 Jurisdictions, consisting of Central Banks and authorities with responsibility of banking regulation. It provides a forum for regular cooperation on banking supervisory matters. Its objective is to enhance understanding of key supervisory issues and improve the quality of banking supervision worldwide. The Committee frames guidelines and standards in different areas – some of the better known among them are the international standards on capital adequacy, the Core Principles for Effective Banking Supervision and the Concordat on cross-border banking supervision. The Committee's Secretariat is located at the Bank for International Settlements (BIS) in Basel, Switzerland. The Bank for International Settlements (BIS) hosts and supports a number of international institutions engaged in standard setting and financial stability, one of which is BCBS. Yet like the other committees, BCBS has its own

governance arrangements, reporting lines and agendas, guided by the central bank governors of the Group of Ten (G10) countries.

37. Basic Balance

The sum of the current-account balance and the net long-term capital flows recorded in a nation's balance of payments accounts.

38. Basic IRR Rule

This is one of the capital budgeting decision rules. Accept the project if IRR (internal rate of return) is greater than the discount rate; reject the project if IRR is less than the discount rate. [See also **Internal rate of return**]

39. Basic Research

A high-risk/high-reward pursuit component of the research and development (R&D) portfolio. Basic research is research to gain knowledge for its own sake.

40. Basic Swap

A plain vanilla interest rate swap in which one party pays a fixed interest rate and receives a floating rate, while the other party pays a floating rate and receives a fixed rate with all rates applied to the same, constant notional principal amount.

41. Basis

The difference between futures price and the spot price. Basis is one kind of risk in investments.

42. Basis Point

When used to describe an interest rate, a basis point is one hundredth of 1% (= 0.01%).

43. Basis Risk

The possibility of unexpected changes in the difference between the price of an asset and the price of the contract hedging the asset. It's the uncertainty that the futures rate minus the cash rate will vary from that expected.

44. Basis Swap

Exchange of floating rate payments between counterparties, but with interest rates based on the different indexes.

45. Basket Credit Default Swap

Credit default swap where there are several reference entities.

46. Basket Option

An option that provides a payoff dependent on the value of a portfolio of assets.

47. Baumol's Economic Order Quantity Model

The Baumol's model strives to equate the two opposing marginal costs associated with ordering and holding inventory to minimize total costs. Just as an operations manager sets inventory levels for raw materials and components, a financial manager can treat cash as a manageable inventory and try to minimize the sum of the following costs:

1. Ordering or transaction costs from converting securities into cash, and vice versa;
2. Carrying or opportunity costs equal to the rate of return foregone to hold cash.

The total costs of cash balances can be defined as:

$$\begin{aligned}\text{Total costs} &= \text{Holding costs} + \text{Transactions costs} \\ &= \frac{C}{2}r + \frac{T}{C}F,\end{aligned}$$

where:

C = Amount of cash raised by selling marketable securities or borrowing

$C/2$ = Average cash balance

r = Opportunity cost of holding cash (the foregone rate of return on marketable securities)

T = Total amount of new cash needed for transaction over entire period (usually 1 year)

T/C = Number of transactions

F = Fixed cost of making a securities trade or borrowing money

The minimum total costs are obtained when C is set equal to C^* , the optimal cash balance. C^* is defined as follows:

$$C^* = \sqrt{\frac{2FT}{r}},$$

where C^* = Optimal amount of cash to be raised by selling marketable securities or by borrowing.

The prior equation represents Baumol's economic order quantity (EOQ) model for determining optimal cash balances. The optimal average cash balance can be defined as $C^*/2$.

In applying Baumol's EOQ model to find an optimal cash balance, the manager must be aware of its underlying assumptions about cash flows:

1. Cash outflows occur at a constant rate.
2. Cash inflows occur periodically when securities are liquidated.
3. Net cash flows also occur at a predictable rate.

EOQ is positively related to F and T and inversely related to r . By taking the square root of FT/r , the relationship with EOQ is less than proportionate. If the value of fixed transaction costs doubles, the EOQ will increase by only 1.41 times.

While the EOQ model offers useful insight into the determination of optimal cash balances, its restrictive assumptions about cash flow behavior are not particularly realistic. Most firms' cash inflows are interspersed with cash outflows. Inflows occasionally exceed flows of outgoing payments. Thus, cash balances over a planning period will move both upward and downward at varying intervals, whereas the EOQ model implicitly assumes demand for cash (inflows) to be positive. Another problem with the EOQ framework is its assumption that inflows (revenue and security sales) are nonrandom and controllable, while outflows (operating costs) are random and uncontrollable. In actuality, control over inflows and outflows is seldom absolute.

48. Bayesian Analysis

A method of statistical inference that allows one to combine prior information about a population parameter with evidence from information contained in a sample to guide the statistical inference process. A prior probability distribution for a parameter of interest is specified first. Recently, this kind of analysis has been used to re-investigate the capital asset pricing model. In addition, Bayesian analysis has frequently been used to revise the beta coefficient for a company.

49. Bear CD

A bear CD pays the holder a fraction of any fall in a given market index.

50. Bear Spread

Bear spread is also called short vertical spread. It is simply the reverse of a long vertical spread. Under this case an investor buys a high-exercise-price call (or put) and sells a low-exercise-price call (or put), both having the same time to expiration left.

51. Bearer Bond

A bond issued without record of the owner's name. Whoever holds the bond (the bearer) is the owner. There are two drawbacks to bearer bonds, first, they can be easily lost or stolen. Second, because the company does not know who owns its bonds, it cannot notify bondholders of important events.

52. Behavioral Finance

Behavioral finance, a subfield of behavioral economics, proposes that psychological influences and biases affect the financial behaviors of investors and financial practitioners. Moreover, influences and biases can be the source for explanation of all types of market anomalies and specifically market anomalies in the stock market, such as severe rises or falls in stock price.

53. Benchmark Analysis

Financial ratios can be used in *benchmark analysis*, in which the ratios of a specific firm can be compared to a benchmark, such as the industry average or an ideal target or goal determined by management. Data for industry average financial ratios are published by a number of organizations, such as Dun and Bradstreet, Robert Morris Associates, Financial Dynamics, Standard and Poor's, and the Federal Trade Commission. These information sources are readily available at most libraries.

54. Benchmark Error

Use of an inappropriate proxy for the true market portfolio.

55. Benchmark Rate

The key driver rate used in sensitivity analysis or simulation models to assess interest rate risk. Other model rates are linked to the benchmark rate in terms of how they change when the benchmark rate changes.

56. Beneficiary

The recipient of the balance in a trust account upon termination of the trust.

57. Benefit/Cost Ratio

A discounted cash flow technique for evaluating capital budgeting projects; more frequently called the **profitability index (PI)**. [See also **Profitability index**]

58. Bermudan Option

An option that can be exercised on specified dates during its life. It is also called as Mid-Atlantic or limited exercise options. This kind of option is a hybrid of American and European options. Instead of being exercised any time before maturity as standard American options, they can be exercised only at a discrete time points before maturity.

59. Best Efforts Offering

A best efforts offering is a less common type of IPO issued by a financially weaker, small, or otherwise risky firm. The investment bank agrees to assist in the marketing of the firm's shares using its best effort and skill but only to sell the shares on a commission basis. The bank buys none of the stock and risks none of its own money. Thus, in a best efforts offering, the issuer bears the risk of price fluctuations or low market demand. If all the shares in a best efforts offering cannot be sold, the issuer may cancel the offering and return all the funds it receives to investors.

Investors should view best efforts offerings with caution. If the knowledgeable investment bank is not willing to risk money to underwrite the firm, why should the investor risk money on the shares.

60. Best-Efforts Underwriting

The underwriter of securities commits to selling as many securities as possible and returns all unsold shares or units to the issuer. [See also **Best efforts offering**]

61. Beta

An estimate of the systematic or market risk of an asset within the capital asset pricing model (CAPM) framework. [See **Beta coefficient**]

62. Beta or Beta Coefficient

We call an asset's or portfolio's systematic risk its beta, denoted by a capital Greek letter β . Beta measures how an asset's returns (R_i) vary with the market portfolio's returns (R_m), compared to the total risk of the market portfolio:

$$\beta_i = \frac{\text{cov}(R_i, R_m)}{\sigma_m^2}$$

where $\text{cov}(R_i, R_m)$ = Covariance between R_i and R_m , σ_m^2 = variance of R_m .

From this perspective, assets that add to portfolio systematic risk will have a high covariance with the market's returns and will therefore have a large beta. Assets that reduce portfolio systematic risk will have a low covariance and a low beta.

Beta is also equal to:

$$\beta = \frac{\rho_{im}\sigma_i\sigma_m}{\sigma_m^2}$$

where σ_i and σ_m are standard deviation of R_i and R_m , respectively, ρ_{im} is the correlation coefficient between R_i and R_m .

If any asset or portfolio has the same exposure to systematic risk as the market portfolio, its beta equals one. Thus, unlike portfolio variance, beta is not an absolute measure of risk. Rather, beta is a measure of relative risk. Beta measures the volatility or variability of an asset's returns *relative* to the market portfolio.

Yet another way to estimate beta is to use regression analysis. To determine an asset's beta, we need to estimate the following regression equation, called the *market model*:

$$R_{it} = \alpha_i + \beta_i R_{mt} + e_{it}$$

where:

R_{it} = the return on the i th asset at time t

R_{mt} = the market return at time t

α_i = the intercept term of the regression

β_i = the slope coefficient of the regression

e_{it} = a random error term

The estimate of the slope is an estimate of the asset's beta since the slope coefficient measures how volatile an asset's returns are relative to the market's returns. If an asset's returns generally rise (or fall) half as much as those of the market, its beta will be 0.5. Knowing that beta equals 0.5 tells us little about the asset's variance of returns over time, or the asset's expected range of returns. As a relative measure of volatility, beta tells us only how, on average, the asset's returns follow those of the overall market.

Assets that are more volatile than the market or, in other words, assets that are more sensitive to systematic risk than the market, have betas greater than 1.0; whatever the market return is, these assets' average returns are larger in *absolute value*. Assets that are less volatile than the market, or those that have less systematic risk, have betas less than 1.0. These assets' returns, on average, are less in absolute value than those of the market.

63. Betting Against Beta

The basic bet against beta strategy is to find assets with higher betas and take a short position in them. At the same time, a leveraged long position is taken in assets with lower betas. The idea is the higher beta assets are overpriced and the lower beta assets are underpriced. The theory posits the prices of the stocks eventually come

back into line with each other. This is essentially a statistical arbitrage strategy with the prices of the assets coming back to the median price versus risk. This median is defined as the security market line (SML).

64. Bid

An offer to purchase at a specified price.

65. Bid Price

The price at which a dealer is willing to purchase a security.

66. Bid-Ask Spread

The difference between a dealer's bid and asked price. The bid is the highest price anyone has declared that he wants to pay for a security at a given time; ask is the lowest price anyone will take at time same time.

67. Bid-Offer Spread

The amount by which the offer price exceeds the bid price. It is also called bid-ask spread. [See also **Bid-ask spread**]

68. Bidder

A firm or person that has made an offer to take over another firm. The bidder offers to pay cash or securities to obtain the stock or assets of another company.

69. BIF

Bank Insurance Fund that insures deposits at commercial banks. This is of the two insurance funds under Federal deposit insurance company (FDIC). [See **SAIF**]

70. Bill of Exchange

A written order requiring a person, business firm, or bank to pay a specified sum of money under stipulated conditions to the bearer of the bill.

71. Bill of Lading

The bill of lading (B/L) is a shipping document that governs transportation of the shipper. Essentially it is a shipping document that governs transportation of the exporter's goods to the importer. The seller submits the invoices and the bill of lading to the correspondent bank. The bank, in turn, verifies the paperwork and pays the seller. The correspondent bank then sends the paperwork to the buyer's bank,

which pays the correspondent bank and sends the documents to the buyer, who makes the payment.

72. Binary Option

Option with a discontinuous payoff; for example, a cash-or-nothing option or an asset-or-nothing option.

73. Binomial Option-Pricing Model

A model where the price of an asset is monitored over successive short periods of time. In each short period, it is assumed that only two price movements are possible. The binomial option pricing model is the most famous binomial model in finance (see Rendelman and Bartter 1979).

74. Binomial Process

[See **Binomial tree**]

75. Binomial Tree

A representation of possible asset price movements over time, in which the asset price is modeled as moving up or down by a given amount each period. This is a special case of a decision tree analysis. [See **Decision trees**]

76. Bivariate Normal Distribution

A bivariate normal distribution is made up of either two independent or two correlated random variables. The two variables in a bivariate normal are both normally distributed, and they have a normal distribution when both are added together. Visually, the bivariate normal distribution is a three-dimensional bell curve. This distribution can be used to derive American options with one dividend payment.

77. Black–Litterman Model

The Black–Litterman (BL) model is an analytical tool used by portfolio managers to optimize asset allocation within an investor's risk tolerance and market views. It then undergoes a process of mean-variance optimization (MVO) to maximize expected return given one's objective risk tolerance.

78. Black-Scholes Option Pricing Model

Black and Scholes came up with a mathematical model to determine the value of an option.

Step 1: The future stock price is constant over time.

Following the equation, $C = \text{Max}(0, S - X)$, where Max denotes the larger of the two bracketed terms, if the stock price is constant over time, then the value of the call, C , is the current price of the stock, S , less the present value of the exercise price, X . Mathematically, the value of the call option, assuming discrete compounding of interest, is:

$$C = S - \frac{X}{(1 + r)^t}$$

If continuous compounding is assumed, then the equation becomes: $V_C = P - Xe^{-rt}$, where e is a constant approximately equal to 2.71828.

Step 2: Assume the price of the stock fluctuates over time.

In this case, we need to adjust the equation for the fluctuation associated with the uncertainty. If we assume that the stock's returns follow a normal distribution, then both S and X in the equation can be adjusted for the uncertainty factor associated with the fluctuation of the stock's price over time. The call option pricing model thus becomes:

$$C = SN(d_1) - Xe^{-rt}N(d_2),$$

where:

$$d_1 = \frac{\ln(S/X) + (r + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}$$

$$d_2 = d_1 - (\sigma\sqrt{t})$$

r = risk-free interest rate

t = time until the option expires (in years)

This equation is the well-known Black-Scholes option pricing model. The adjustment factors $N(d_1)$ and $N(d_2)$ represent the cumulative standard normal distribution function. $N(d_1)$ and $N(d_2)$ are probabilities that a random variable with a standard normal distribution takes on a value less than d_1 and d_2 , respectively. The values for $N(d_1)$ and $N(d_2)$ can be found by using a standardized normal distribution table.

79. Black's Model (Formula)

A version of Black-Scholes formula in which the underlying asset is a futures price and the dividend yield is replaced with the risk-free rate. The formula is written as follows:

$$C(F, K, \sigma, r, t, r) = Fe^{-rt}N(d_1) - Ke^{-rt}N(d_2),$$

$$d_1 = \frac{\ln(F/K) + \frac{1}{2}\sigma^2 t}{\sigma\sqrt{t}},$$

$$d_2 = d_1 - \sigma\sqrt{t},$$

where F = futures price; K = exercise price; r = risk-free rate; σ = standard deviation of rates of return; and t = contract period.

The put price is obtained using the parity relationship for options of futures:

$$P(F, K, \sigma, r, t, r) = C(F, K, \sigma, r, t, r) + Ke^{-rt} - Fe^{-rt}$$

80. Black-Scholes Formula

An equation to value a call option that uses the stock price, the exercise price, the risk-free interest rate, the time to maturity, and the standard deviation of the stock return.

$$C = SN(d_1) - Xe^{-rt}N(d_2)$$

where

$$d_1 = \frac{\ln(S/X) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

C = Current call option value

S = Current stock price

$N(d)$ = The probability that a random draw from a standard normal distribution will be less than d ; equals the area under the normal curve up to d

X = Exercise price

e = 2.71828, the base of the natural log function

r = Risk-free interest rate

$\ln$ = Natural logarithm function

σ = Standard deviation of the annualized continuously compounded rate of return of the stock

Like most of the models, the Black-Scholes formula is based on some important underlying assumptions:

1. The stock will pay no dividends until after the option expiration date.
2. Both the interest rate, r , and variance rate, σ^2 , of the stock are constant.

3. Stock prices are continuous, meaning that sudden extreme jumps such as those in the aftermath of an announcement of a takeover attempt are ruled out.

81. Blank Check

A signed check with no amount indicated.

82. Blanket Lien

A loan may specify a blanket lien, or a claim, against all work in progress or inventory on hand. For example, if a business mass produces low-value items, it is not practical to give the bank claims against specific items. In other words it refers a secured loan that gives the lender a lien against all the borrower's inventories. [See also **Collateral**]

83. Blanket Mortgage

A blanket mortgage is a claim on all the issuer's real property, including land, buildings, and equipment.

84. Block House

Brokerage firms that help to find potential buyers or sellers of large block trades.

85. Block Sale

A transaction of more than 10,000 shares of stock.

86. Block Transactions

Large transactions in which at least 10,000 shares of stock are bought or sold. Brokers or "block houses" often search directly for other large traders rather than bringing the trade to the stock exchange.

87. Board Broker

Individual who handles limit orders in some exchanges. The board broker makes information on outstanding limit orders available to other traders.

88. Board of Directors

Individuals elected by stockholders to manage and oversee a firm's operations.

89. Board of Governors of the Federal Reserve System

The policy-setting representatives of the Federal Reserve System in charge of setting the discount rate, required reserves, and general policies designed to affect growth in the banking system's reserves and US money supply.

90. Bogey

In portfolio performance analysis, the attribution method explains the difference in returns between a managed portfolio and selected benchmark portfolio is called the bogey. Attribution studies start from the broadest assets allocation choices and progressively focus on ever-finer details of portfolio choice.

91. Bond

A long-term debt of a firm. In common usage, the term *bond* often refers to both secured and unsecured debt. Bonds usually have a face value, it is also called the principal value or the denomination and it is stated on the bond certificate. In addition, the par value (i.e., initial accounting value) of a bond is almost always the same as the face value.

92. Bond Anticipation Notes

Shorter-term securities issued by a state or local government to raise funds to begin a project that eventually will be funded by issuing long-term bonds.

93. Bond Broker

A broker who trades bonds on an exchange.

94. Bond-Equivalent Basis

With bond-equivalent basis, yield calculations use the same number of days for both interest-bearing periods and interest-compounding periods. For example, to figure the annual yield using daily compounding, the annual interest rate might be divided by 365 days, and the result then compounded for 365 days to get the annual yield.

95. Bond Equivalent Yield of the T-Bill

Bond equivalent yield of T-bill (Y_{BEY}) is generally calculated on an annual percentage rate (APR) method as follows.

$$Y_{\text{BEY}} = \frac{10,000 - P}{P} \times \frac{365}{n}$$

where P and n represent market price and the number of days owned the T-bill. It differs from the effective annual yield method. [See **APR, Effective annual yield**]

96. Bond Fund

A mutual fund that invests in debt instruments. It is an income fund instead of growth fund.

97. Bond Option

An option where a bond is the underlying asset.

98. Bond Ratings

Most issuers secure bond ratings from one or more agencies such as Standard and Poor (S&P), Moody's, Fitch, and Duff and Phelps. From its analysis and discussions with management, the agency assigns a bond rating. In addition, the rating agency commits to a continual reexamination of the issue's risk. For example, should the financial position of the firm weaken or improve, S&P may place the issue on its *Credit Watch* list, with negative or positive implications. Shortly thereafter, S&P will downgrade, upgrade, or reaffirm the original rating.

Bond rating is another example of an agency cost. [See also **Agency costs**] To show potential investors the credit quality of its bonds, the firm hires a recognized independent third party to rate its bond offering. Even if the bonds receive a lower-than-expected rating and the firm must issue the bonds with a higher coupon to compensate investors for the extra risk, the benefits of a rating in terms of attractiveness to investors and issue liquidity outweigh these extra costs.

Despite the initial cost and the concern that a lower-than-expected rating may cause managers, a bond rating makes it much easier to sell the bonds in the primary market offering, as well as in the secondary market. The rating acts as a signal to the market that an independent agency has examined the qualities of the issuer and the issue and has determined that the credit risk of the bond issue justifies the published rating. An unrated bond issue risks a cool reception in the primary market and thin illiquid, secondary markets (i.e., bond traders and investors are not interested in buying or selling the particular issues). Investors may have good reason to wonder, "What is the firm trying to hide? If this really was an attractive bond issue, the firm would have had it rated." In addition, certain types of investors, such as pension funds and insurance companies, may face restrictions against purchasing unrated public debt. Examples of bond rating categories are presented in Table 3.

Table 3 Alternative bond rating categories

	Moody's	Standard and Poor's	Former Standard and Poor's
Best quality, smallest degree of risk	Aaa	AAA	AAA
High quality, slightly more long-term risk than top rating	Aa1	AA+	AA
	Aa2	AA	
	Aa3	AA—	
Upper-medium grade, possible impairment in the future	A1	A+	A
	A2	A	
	A3	A—	
Medium-grade, lack outstanding investment characteristics	Baa1	BBB+	BBB
	Baa2	BBB	
	Baa3	BBB—	
Speculative issues, protection may be very moderate	Ba1	BB+	BB
	Ba2	BB	
	Ba3	BB—	
Very speculative, may have small assurance of interest and principal payments	B1	B+	B
	B2	B	
	B3	B—	
Issues in poor standing, may be in default	Caa	CCC	CCC
Speculative in a high degree, with marked shortcomings	Ca	CC	CC
Lowest quality, poor prospects of attaining real investment standing	C	C	C
		D	

99. Bond Strategies

There are multiple options available when it comes to structuring a bond portfolio, and each strategy comes with its own risk and reward trade-offs. The four principal strategies used to manage bond portfolios are Passive or “buy and hold,” Index matching or “quasi-passive,” Immunization or “quasi-active,” and Dedicated and active. The buy-and-hold approach appeals to investors who are looking for income and are not willing to make predictions. The middle-of-the-road strategies include indexation and immunization, both of which offer some security and predictability. Then there is the active world, which is not for the casual investor. Each strategy has its place and when implemented correctly, can achieve the goals for which it was intended. Some well-known bond strategies are (i) writing the yield curve, (ii) maturity-structure strategies, (iii) swapping, (iv) inter-market spread swap, (v) interest rate anticipation swap, and (vi) pure-yield pickup swap.

100. Bond Valuation

Theoretical value of bond is equal to present value of annuity for future interest payments and present value of the face value of bond. Suppose that a bond with par (face) value F is purchased today and that the bond matures in N years. Let us assume

that interest payments of dollar amount I are to be made at the end of each of the next N years. The bondholders will then receive a stream of N annual payments of I dollars, plus a payment of F dollars at the end of the N th year. Using the rate of interest r to discount future receipts, the present value of the bond is:

$$PV = \sum_{t=1}^N \frac{I}{(1+r)^t} + \frac{F}{(1+r)^N} \quad (2)$$

The first term on the right-hand side of Eq. 2 is the present value of the stream of interest payments, while the second term is the present value of the future of the par amount.

101. Bond Yield

Discount rate which, when applied to all the cash flows of a bond, causes the present value of the cash flows to equal the bond's market price. [See **Yield to maturity**]

102. Book Cash

A firm's cash balance as reported in its financial statements. Also called ledger cash. It is not the same thing as the balance shown in its bank account (bank cash or collected bank cash).

103. Book Value

Under US GAAP, balance sheet items generally are listed at book value, which is the original or historical cost of the items, less calculated depreciation. Frequently, book value fails to accurately represent the current market value of balance sheet items. For example, LIFO (last-in, first-out) inventory accounting may leave items produced years ago on the books even though they were actually sold time ago. On the other hand, last year's fashions or models may be nearly worthless in the current market, while the balance sheet inventory figure values them at their historical cost. Historical cost may grossly undervalue fixed assets such as land or buildings. Likewise, bond issues, valued at par depending upon interest rate or credit risk changes since they were first issued. Similarly, book values of equity claims will differ from current market prices of company stock issues.

104. Book Value Equity

Total assets minus total liabilities reported on the balance sheet. It includes the par value, capital surplus and accumulated retained earnings.

105. Book Value Per Share

Per-share accounting equity value of a firm. Total accounting equity divided by the number of outstanding shares. The sum of the par value, capital surplus, and

accumulated retained earnings is the common equity of the firm, which is usually referred to as the firm's book value. The book value represents the amount contributed directly and indirectly to the corporation by equity investors.

106. Book-Entry Form

The method by which marketable US Treasury securities are issued today with the buyer receiving only a receipt, rather than an engraved certificate, indicating that the purchase is recorded on the Treasury's books or recorded in another approved location.

107. Book-Tax Differences

Book income describes a company's financial income before taxes. It is the amount a corporation reports to its investors or shareholders and gives an idea of how well a company performed during a certain period of time. Tax income, on the other hand, is the amount of taxable income a company reports on its return. The purpose of the Schedule M-1 is to reconcile the entity's accounting income (book income) with its taxable income. Because tax law is generally different from book reporting requirements, book income can differ from taxable income.

108. Bootstrapping

This term has two meanings. First, it refers to the procedure where coupon bonds are used to generate the set of zero-coupon bond prices. Second, it means the use of historical returns to create an empirical probability distribution for returns.

109. Borrow

To obtain or receive money on loan with the promise or understanding of returning it or its equivalent.

110. Borrowed Reserves

Legal reserves loaned to depository institutions through the discount windows of the Federal Reserve banks.

111. Borrowing

The change in liabilities outstanding reported by a sector or unit in the economy over a specified time period.

112. Borrowing Portfolio

In the capital market line, assume that the investor can borrow money at the risk-free rate and invest the money in the risky portfolio. The portfolios with a rate of

return higher than the return on market portfolio, but with higher risks along the capital market line contains a negative amount of the risk-free asset is called borrowing portfolios. The negative amount invested in the risk-free asset can be viewed as borrowing funds at the risk-free rate and investing in risky assets.

113. Bounce a Check

A depositor writes a check which is returned to the bank and by the bank to the depositor because of insufficient funds.

114. Boundary Condition

The value of a derivative claim at a certain time, or at a particular price of the underlying asset. For example, a boundary condition for a zero-coupon bond is that the bond at maturity is worth its promised maturity value.

115. Box Spread

An option position in which the stock is synthetically purchased (buy call, sell put) at one price and sold (sell call, buy put) at a different price. When constructed with European options, the box spread is equivalent to a zero-coupon bond.

116. Branch Banking

An organizational structure in which a bank maintains facilities that are part of the bank in offices different from its home office. Some states allow banks to set up branches through the state, county, or city. Others prohibit branches.

117. Break-Even Analysis

Analysis of the level of sales at which a project would make zero profit. We calculate the break-even point in terms of both accounting profit and present value. [See **Break-even point**, **Accounting break-even point**, **Cash break-even point**, and **Financial break-even point**]

118. Break-Even Point

A firm's break-even point is where revenues equal total costs. [See also **Accounting break-even point** and **Cash break-even point**] Break-even point (Q^*) can be defined as:

$$Q^* = \frac{F + D}{P - V}$$

where F , P , V , and D represent total fixed cost, price per unit, variable cost per unit, and depreciation, respectively.

119. Break Point

A break point occurs when raising an additional dollar of funds results in an increase in the weighted average cost of capital.

We know that firms have two sources of equity financing, each with a different cost: the cost of retained earnings (k_{re}) and the cost of new common stock (k_{cs}). In financing a number of attractive projects, the firm may deplete its retained profits from the current year. A need for additional financing creates a need to recalculate a cost of capital that substitutes the cost of new common stock for the cost of retained earnings. The firm must discount additional investments using the new, incremental or *marginal weighed average cost of capital* as a discount rate.

Similarly, a firm may be able to borrow only a limited amount in a year without harming its credit or bond rating. A downgrade would increase the interest rates it pays to borrow funds. Once the firm reaches its debt limit, it must calculate a new cost of capital that incorporates the higher cost of borrowing. Any additional investment then must face a higher cost of capital because of the increase in borrowing costs.

Thus, several break points may exist for a firm. One may arise from depleting current retained profits; another may arise from higher borrowing costs if substantial funds are borrowed in a given year; still others may arise from changes in flotation costs or equity costs should large amounts be raised from these sources in a short time frame.

120. Bridge Loan

A loan issued to fund a temporary need from the time a security is redeemed to the time another security is issued.

121. Broker

An individual who executes orders for customers for which he/she receives a commission.

122. Brokered Deposit

Deposits acquired through a money broker (typically an investment bank) in the national markets.

123. Brokered Market

A market where an intermediary (a broker) offers search services to buyers and sellers.

124. Brownian Motion

A stochastic process in which the random variable moves continuously and follows a random walk with normally distributed independent increments. Named after the Scottish botanist Robert Brown, who in 1827 noticed that pollen grains suspended in water exhibited continual movement. Brownian motion is also called a *Wiener process*. This is a basic concept used to derive the continuous type of option pricing model. [See also **Wiener Process**]

125. Bubble Theory (of Speculative Markets)

Bubble refers to security price move wildly above their true values and eventually burst. After prices eventually fall back to their original level, causing great losses for investors. The crashes of stock markets of United States in 1929, 1987, and 2000 are evidences for the bubble theory.

126. Budget Deficit

The amount by which government spending exceeds government revenues. Fiscal condition for a government can be either budget deficit or budget surplus.

127. Budget Surplus

A government's financial position in which current revenues exceed current expenditures.

128. Bulge Bracket Firms

Firms in an underwriting syndicate that has the highest commitment to assist in placing the underlying securities.

129. Bull CD

A bull CD pays its holder a specified percentage of the increase in return on a specified market index while guaranteeing a minimum rate of return.

130. Bull Spread

Bull spread is also called long vertical spread. It designates a position for which one has brought a low-exercise-price call (or a low-exercise-price put) and sold a high-exercise-price call (or a high-exercise price put) that both mature in the same month.

131. Bullet Loan

A loan that requires payment of the entire principal at maturity.

132. Bullish, Bearish

Words used to describe investor attitudes. *Bullish* means optimistic; *bearish* means pessimistic. Also used in bull market and bear market for describing the stock market.

133. Bundling, Unbundling

A trend allowing creating of securities either by combining primitive and derivative securities into one composite hybrid or by separating returns on an asset into classes. Both cases are financial engineering techniques. Creative security design often calls for bundling primitive and derivative securities into one composite security. Quite often, creating a security that appears to be attractive requires unbundling of an asset. A mortgage pass-through certificate is unbundled into two classes. Class 1 receives only principal payments from the mortgage pool, whereas class 2 receives only interest payments.

134. Burden

Noninterest expense minus noninterest income for banks. Generally, noninterest expense is larger than the noninterest income.

135. Business Cycle

Repetitive cycles of recession and recovery. Some cyclical indicators for business cycle based upon National Bureau of Economic Research are as follows:

(a) Leading Indicators

- Average hourly workweek, production workers, manufacturing
- Average weekly initial claims, state unemployment insurance
- Index of net business formation
- New orders, durable-goods industries
- Contracts and orders, plant and equipment
- Index of new building permits, private housing units
- Change in book value, manufacturing and trade inventories
- Index of industrial materials prices
- Index of stock prices, 500 common stocks
- Corporate profits after taxes (quarterly)
- Index: ratio of price to unit labor cost, manufacturing
- Change in consumer installment debt

(b) Roughly Coincident Indicators

- GNP in current dollars
- GNP in 1958 dollars
- Index of industrial production
- Personal income
- Manufacturing and trade sales
- Sales of retail stores
- Employees on nonagricultural payrolls
- Unemployment rate, total

(c) Lagging Indicators

- Unemployment rate, persons unemployed 15 weeks or over
- Business expenditures, new plant and equipment
- Book value, manufacturing and trade inventories
- Index of labor cost per unit of output in manufacturing
- Commercial and industrial loans outstanding in large commercial banks
- Banks rates on short-term business loans

Source: US Department of Commerce.

136. Business Failure

It refers to a business that has terminated due to the loss of creditors. However, it should be noted that even an all-equity firm can fail.

137. Business Risk

Business risk is determined by the products the firm sells and the production processes it uses. The effects of business risk are seen ultimately in the variability of earnings before interest and taxes (EBIT) over time. In fact, one popular measure of a firm's business risk is the standard deviation of EBIT. To control for the effects of a firm's size, another popular method of gauging business risk is to find the standard deviation over time of the firm's operating return on assets, that is, EBIT divided by Total assets.

A firm's business risk is affected by three major influences: unit volume fluctuations, fixed costs (including depreciation expenses), and the relationship between the firm's selling price and its variable costs.

138. Business Strategy Matrix

The business strategy matrix model views the firm as a collection or portfolio of assets grouped into strategic business units. This technique has been disparaged by some as a cause of inappropriate diversification among business units. It has led firms to acquire or develop unrelated business units that the firm's officers did not fully understand. For example, the managerial expertise needed to run a successful

electronics firm may be different from that needed to run a successful baking company. Nonetheless, this model can still provide some insights into capital budgeting strategy.

The business strategy matrix model emphasizes market share and market growth rate. Based upon these attributes, business units are deemed to be Stars, Cash Cows, Question Marks, or Dogs. Cash Cows typically are business units with leading market positions in maturing industries; the firm can direct the cash that these units generate to other business units that need it, such as Stars and Question Marks. The Stars (units with good market positions in high growth markets) need funds to expand and develop competitive advantages, as do some Question Marks. Proper strategies to build competitive advantages may turn Question Marks into Stars; if these strategies are unsuccessful, the firm may have to divest Question Marks. Dogs have poor market positions in low-growth industries; unless a turn-around strategy is feasible, these also are divestment or liquidation candidates.

If an organization uses the business strategy matrix to assist in planning, management must be sure to manage the firm's market share in a way that maximizes shareholder wealth.

139. Butterfly Spread

A position that is created by taking a long position in a call with strike price K_1 , a long position in a call with strike price K_3 , and a short position in two calls with strike price K_2 , where $K_3 > K_2 > K_1$ and $K_2 = 0.5(K_1 + K_3)$. (A butterfly spread can also be created with put options.)

140. Buying the Index

Purchasing the stocks in the Standard and Poor's 500 in the same proportion as the index to achieve the same return. An index fund is a good example.

1.3 C

1. Cable Transfers

Orders transferred by a cable or wire to a foreign bank holding the account of a seller of currency directing that a stipulated amount of a particular currency be moved into another account.

2. Calendar Spread

A position that is created by taking a long position in a call option that matures at one time and a short position at a different time. (A calendar spread can also be created using put options.)

3. Calibration

Method for implying a model's parameters from the prices of actively traded options.

4. Call

A call is an option to purchase a fixed number of shares of common stock. It is a right instead of an obligation. Call can be either buy a call or write a call.

5. Call Deferment Periods

[See **Callable bonds**]

6. Call Loan

A call loan is a loan contract which enables the lender (e.g., the bank) to request repayment of loan in the contract period. For example, most broker loans to investment banks are callable within 24 h notice.

7. Call Money Rate

It is the rate charged by brokers for the use of margin in common-stock accounts.

8. Call Option

A call option gives the holder the right to buy a particular number of shares of a designated common stock at a specified price, called the **exercise price** (or striking price), on or before a given date, known as the **expiration date**. [See also **Exercise price** and **Expiration date**] On the Chicago Board Options Exchange, options typically are created for 3-, 6-, or 9-month periods. All have the same expiration date: the Saturday following the third Friday of the month of expiration. The owner of the shares of common stock can write, or create, an option and sell it in the options market, in an attempt to increase the return or income on a stock investment.

9. Call Premium

It refers to the price of a call option on common stock. It also refers to the difference between the call price and the face. [See also **Callable bonds**]

10. Call Price of a Bond

Amount at which a firm has the right to repurchase its bonds or debentures before the stated maturity date. The call price is always set at equal to or more than the par value. [See **Callable bonds**]

11. Call Privilege

The provision often found in a bond's contract (indenture) that permits the borrower to retire all or a portion of a bond issue by buying back the securities in advance of their maturity.

12. Call Protected

Described a bond that is not allowed to be called, usually for a certain early period in the life of the bond.

13. Call Protection

The feature which does not allow a bond to be called for some (deferment) period.

14. Call Provision

A written agreement between an issuing corporation and its bondholders that gives the corporation the option to redeem the bond at a specified price before the maturity date. A call provision lets the company repurchase or call the entire bond issue at predetermined price over a specific period.

15. Call Risk

[See **Callable bonds**]

16. Call-Loan Money Rate

The rate charged by banks to brokers who deposit securities as collateral.

17. Callable

Refers to a bond that is subject to be repurchased at a stated call price before maturity. For example, debt may be extinguished before maturity by a call. Historically, almost all publicly issued corporate long-term debt has been callable.

18. Callable Bonds

Callable bonds can be redeemed prior to maturity by the firm. Such bonds will be called and redeemed if, for example, a decline in interest rates makes it attractive for the firm to issue lower coupon debt to replace high-coupon debt. A firm with cash from successful marketing efforts or a recent stock issue also may decide to retire its callable debt.

Callable bonds usually are called away after a decline in interest rates. As rates fall, the bond's price will not rise above its call price. Thus, for callable bonds, the

inverse relationship between bond prices and interest rates breaks down once the bond's market price reaches the call price.

Many indentures state that, if called, callable bonds must be redeemed at their **call prices**, typically par value plus a **call premium** of 1 year's interest.

Investors in callable bonds are said to be subject to **call risk**. Despite receiving the call price, investors usually are not pleased when their bonds are called away. As bonds typically are called after a substantial decline in interest rates, the call eliminates their high coupon payments; they can reinvest the funds only in bonds that offer lower yields.

In order to attract investors, callable bonds must offer higher coupons or yields than noncallable bonds of similar credit quality and maturity. Many indentures specify **call deferment periods** immediately after the bond issue, during which the bonds cannot be called.

19. CAMELS

An acronym that refers to the regulatory rating system for bank performance: C = capital adequacy, A = asset quality, M = management quality, E = earnings quality, L = liquidity, and S = sensitivity to market risk.

20. Cancelable Swap

A cancelable swap is a plain vanilla interest rate swap. This kind of swap can be cancelled by one side on prespecified dates.

21. Cannibalization

Cannibalization occurs when a project robs cash flow from the firm's existing lives of business. For example, when a soft-drink firm is thinking about introducing a new flavor or a new diet product, the project's incremental cash flows should consider how much the new offering will erode the sales and cash flows of the firm's other product lines.

22. Cap

An options contract that serves as insurance against a high price. [See also **Interest rate cap**]

23. Cap Rate

The rate determining payoffs in an interest rate cap. [See **Interest rate Cap**]

24. Capacity Utilization Index

The capacity utilization rate measures the proportion of potential economic output that is actually realized. Displayed as a percentage, the capacity utilization

level provides insight into the overall slack that is in an economy or a firm at a given point in time. The capacity utilization rate is important for assessing a company's current operating efficiency, and it helps provide insight into cost structure in the short term or long term. It can be used to determine the level at which unit costs rise. In general, asset turnover ratios can be used to approximately measure the capacity utilization index.

25. Capital

Funds subscribed and aid by stockholders representing ownership in a bank. In other word, capital is the stockholder's equity of a bank. Regulatory capital also includes debt components and loss reserves. It can be defined either in book value or market value. The market value of capital is used as insulation device against credit risk and interest rate risk. [See **Credit risk** and **Interest rate risk**]

26. Capital Account

A record of flows of short-term and long-term funds into a nation and out of a nation and included in its balance-of-payments accounts.

27. Capital Allocation Decision

Allocation of invested funds between risk-free assets versus the risky portfolio. [See also **Asset allocation decision**]

28. Capital Allocation Line (CAL)

A graph showing all feasible risk-return combinations of a risky and risk-free asset. [See also **Capital market line**]

29. Capital Asset Pricing Model (CAPM)

An equilibrium asset pricing theory that shows that equilibrium rates of expected return on all risky assets are a function of their covariance with the market portfolio (see Sharpe 1964).

$$E(R_i) = R_F + \beta(E(R_m) - R_F)$$

Thus, expected return on i th security = Risk-free rate + Beta coefficient (Expected return on market portfolio – Risk-free rate). Because the term in parentheses on the right-hand side is positive, this equation says that the expected return on a security is a positive function of its beta.

30. Capital Budgeting

Capital budgeting is the process of identifying, evaluating, and implementing a firm's investment opportunities. Because of their size and time horizon, a firm's

capital projects should reflect its strategy for meeting future goals. The typical capital budgeting project involves a large up-front cash outlay, followed by a series of smaller cash inflows and outflows. A project's expected time frame may be as short as 1 year or as long as 20 or 30 years. But the project's cash flows, including the total upfront cost of the project, are not known with certainty before the project starts. The firm must evaluate the size, timing, and risk of the project's cash flows to determine if it enhances shareholder wealth.

Broadly speaking capital budgeting can be described as a three-phase process that includes a **planning phase**, an **appropriation phase**, and an **audit, or control, phase**. [See also **Planning phase**, **Appropriation phase**, and **Audit, or control, phase**]

31. Capital Budgeting Process

The capital budgeting process includes the identification phase, the development phase, the selection phase, and the control phase.

32. Capital Budgeting Under Certainty

When making investment decisions, we assume that net cash inflow, discount rate, machine life, salvage value, and other factors are known for sure. This kind of decision process is called capital budgeting under certainty. There are five methods of capital budgeting under certainty. The methods are payback method, average accounting rate of return method, net present value method, internal rate of return method, and profitability index method.

33. Capital Budgeting Under Uncertainty

In the real world, the input variables for investment decisions, such as net cash inflow, discount rate, machine life, salvage value, and other factors, are not known for sure. Therefore, we need to use capital budgeting under uncertainty methods, such as the risk-adjusted discount rate method, certainty equivalent method, statistical distribution method, decision tree method, and simulation method, to make investment decisions.

34. Capital Gains

The positive change in the value of an asset. A negative capital gain is a capital loss. It is the change in the price of the stock divided by the initial price. Letting P_t be the purchaser price of the asset and P_{t+1} be the price of the asset at year-end, the capital gain can be computed.

$$\text{Capital gain} = (P_{t+1} - P_t)/P_t$$

35. Capital-Labor Ratio

The capital-labor ratio can measure the capital intensity of a firm. Typically, over time, firms tend to have a higher capital-labor ratio as they seek to gain productivity

improvements from investment in capital and automating the production process. For example, the Cobb-Douglas production function and the CS production function have used both capital and labor in their production functions.

36. Capital Lease

Capital LA capital lease or **financial lease** of an asset satisfies any one of the following criteria:

1. The lessee takes ownership of the asset at the end of the lease.
2. The lessee can purchase the asset at the end of the lease at a bargain price (less than fair market value).
3. The length of the lease equals 75% or more of the estimated life of the asset.
4. At the beginning of the lease, the present value of the lease payments is 90% or more of the fair market value of the property.

Typically, the lessee may not cancel a capital or financial lease and is responsible for asset maintenance. In a financial lease, tax law identifies the lessor as the owner of the leased asset, so the lessor can deduct asset depreciation over the life of the lease.

37. Capital Markets

Financial markets for long-term debt (with a maturity at over 1 year) and for equity shares. The financial markets are composed of the money markets and the capital markets. The markets where capital, such as stocks and bonds, are traded. Capital markets are used by firms to raise additional funds.

38. Capital Market Line

The efficient set of all assets, both risky and riskless, which provides the investor with the best possible opportunities. The line used in the risk-return trade off to illustrate the rates of return for efficient portfolios depending on the risk free rate of return and the level of risk (standard deviation) for a particular portfolio.

In sum, formula for capital market line used to describe the trade-off between expected return and total risk is

$$E(R_i) = R_f + [E(R_m) - R_f] \frac{\sigma_i}{\sigma_m},$$

where:

R_f = risk-free rate

$E(R_m)$ = expected return on the market portfolio

$E(R_i)$ = expected return on the i th portfolio

σ_i, σ_m = standard deviations of the portfolio and the market, respectively

39. Capital Market Securities

The classification of a financial instrument as a marketable security typically is based upon maturity and, to a lesser extent, liquidity. Securities with more than 1 year to maturity, such as stocks, bonds, and mortgages, are called capital market securities.

40. Capital Rationing

Capital rationing places an upper limit on the amount of a firm's capital spending over the course of a year. The first break point in a cost of capital schedule usually occurs when a firm runs out of current retained earnings. An easy way to handle the marginal cost of capital problem is to ration capital by setting the upper limit of spending at the point where the firm will run out of retained earnings.

The deficiency in this strategy is rather obvious. To maximize shareholder wealth, the firm should be willing to undertake *any* project with a positive NPV, whether or not total spending exceeds one or more break points.

41. Capital Structure

The mix of debt and equity a firm uses to finance its assets defines the firm's capital structure. A target capital structure is important as it determines the weights in the calculation of a firm's weighted average cost of capital. There is, however, a second and even more important reason: The firm's *optimum* debt-to-equity mix minimizes the WACC; minimizing the WACC will help the firm to maximize shareholder wealth.

If a nonoptimal capital structure leads to a higher WACC, the firm is likely to reject some capital budgeting projects that could increase its competitive advantage and shareholder wealth under an optimal target financing mix.

42. Capital Structure Ratios

Capital structure ratios (sometimes called *debt utilization* or *leverage ratios*) compare the funds supplied by the owners (equity) with the funds provided by creditors (debt). The **debt-to-assets ratio** is calculated as total debt (i.e., the sum of current and long-term liabilities) divided by total assets; it measures the proportion of assets financed by borrowers. The **debt-to-equity ratio** is computed as total debt divided by stockholders' equity. The two ratios are computed as:

$$\text{Debt-to-assets ratio} = \frac{\text{Total debt}}{\text{Total assets}}$$

$$\text{Debt-to-equity ratio} = \frac{\text{Total debt}}{\text{Total equity}}$$

The **equity multiplier** is another indicator of a company's use of debt. At first glance, the ratio appears to have little to do with leverage; it is simply total assets divided by stockholders' equity. Recall the accounting identity, however:

Assets = Liabilities + Equity. More assets relative to equity suggest a greater use of debt. Thus, larger values of the equity multiplier imply a greater use of leverage by the firm. The equity multiplier is written as:

$$\text{Equity multiplier} = \frac{\text{Total assets}}{\text{Total equity}}$$

As a rough measure of the firm's ability to service its debt and other fixed obligations, the analyst can calculate the **times interest earned** (or **interest coverage) ratio**. The times interest earned (TIE) ratio is earnings before interest and taxes (EBIT) divided by interest expense. This ratio provides a measure of how well the firm's operations generate funds to pay interest expenses. EBIT can fall by $(1-1/\text{TIE})$ before interest payments are jeopardized. For example, a TIE ratio of 5 indicates that EBIT could fall by $(1-1/5)$ or 80% before earnings would fail to cover interest obligations. The times interest earned ratio is given by:

$$\text{Times interest earned ratio} = \frac{\text{EBIT}}{\text{Interest expense}}$$

An alternative to the TIE ratio, the **fixed charge coverage ratio** is computed as earnings before fixed charges divided by fixed charges. It is more general than TIE, since the denominator includes all fixed charges, such as interest payments, lease payments, and bond sinking fund obligations. However, looking at the fixed charge coverage ratio may give analysts a fuller picture of the firm's ability to pay all of its fixed obligations.

By using appropriate amounts of debt and equity, the firm can minimize its financing costs and thereby maximize shareholder wealth. This suggests that analysts may see danger signals in both high and low ratios. High debt ratios increase the potential of bankruptcy; low debt ratios may indicate that management is not using debt efficiently to maximize shareholder wealth.

43. Capital Surplus

Amounts of directly contributed equity capital in excess of the par value. Equity that cannot otherwise be classified as capital stock or retained earnings. It is usually created from a stock issued at a premium over par value. Capital surplus is also known as share premium (UK), acquired surplus, donated surplus, paid-in surplus, or additional paid-in capital.

44. Capital-Labor Ratio

A production function is a function that can be seen as a function of labor and capital:

$$Q = f(K, L)$$

where K = capital and L = labor. K/L (the capital-labor ratio) is generally used to measure a firm's degree of capital intensity. Capital intensity results in increased total risks and generally results in an increase in beta. If the capital-labor ratio is greater than 1 – that is, if K is greater than L – a firm is capital intensive. If the ratio is less than 1, then there is a deduction in capital intensity and a shift toward human-resource investment.

45. Captive Finance Company

A finance company owned by a manufacturer that provides financing to buyers of the firm's products. For example, General Motors Acceptance Corporation is a captive finance company.

46. Caplets

The individual options comprising a cap are sometimes referred to as caplets. An interest rate cap is a series of consecutive long call options (caplets) on a specific interest rate at the same strike price.

47. Capped Option

An option with a maximum payoff, where the option is automatically exercised if the underlying asset reaches the price at which the maximum payoff is attained.

48. CARs

Collateralized automobile receivables- a form of asset-backed security in which the collateral is automobile receivables. Other type of account receivable can be used to create asset-backed security also.

49. Car

A loose term sometimes used to describe the quantity of a contract – for example, “I am long a car of bellies.” (Derived from the fact that quantities of the product specified in a contract used to correspond closely to the capacity of a railroad car.)

50. CARA Utility Function

Constant absolute risk aversion (CARA), or negative exponential utility functions, with an assumed value of ARA often is used to analyze farm decisions under risk. For CARA utility functions, an ARA value is sufficient to determine an individual's preference over alternative decisions. For a given ARA level, one can determine how preferences change as risk aversion increases. However, an ARA value does not convey sufficient information to indicate whether the implied level of risk aversion is “reasonable.” One typically cannot tell from knowledge of the

assumed ARA level if, for example, the preference of policy A over policy B is held by slightly, moderately, or strongly risk-averse individuals or by those with unreasonably high levels of risk aversion.

51. Card Bank

Bank that administers its own credit card plan or serves as a primary regional agent of a national credit card operation.

52. Cardinal Utility

A cardinal utility implies that a consumer is capable of assigning to every commodity or combination of commodities a number representing the amount or degree of utility associated with it.

53. Carry

Another term for owning an asset, typically used to refer to commodities. [See also **Carry market** and **Cost of carry**]

54. Carry Market

A situation where the forward price is such that the return on a cash-and-carry is the risk-free rate. Cash-and-carry refers to the simultaneous spot purchase and forward sale of an asset or commodity

55. Carrying Costs

Costs that increase with increases in the level of investment in current assets. Costs that fall with increases in the level of investment in current assets are called shortage costs. Carrying costs are generally of two types. First, because the rate of return on current assets is low compared with that of other assets, there is an opportunity cost. Second, there is the cost of maintaining the economic value of the item. For example, the cost of warehousing inventory belongs here.

56. Carrying Value

Book value. It is an accounting number based on cost.

57. Carve Outs

[See **Voluntary restructuring**]

58. Cash Basis

The accounting procedure that recognizes revenues when cash is actually received and expenses when cash is actually paid.

59. Cash Budget

A forecast of cash receipts and disbursements expected by a firm in the coming year. It is a short-term financial planning tool. It allows the financial manager to identify short-term financial needs (and opportunities). It will tell the manager the required borrowing for the short term. It is the way of identifying the cash-flow gap on the cash-flow time line. The basic relation is:

$$\begin{aligned} &\text{Ending accounting receivable} \\ &= \text{Starting accounting receivable} \\ &+ \text{Sales} - \text{Collection} \end{aligned}$$

Collection is not the only source of cash; other sources of cash include sales of assets, investment income, and long-term financing.

60. Cash Break-Even

Cash break-even occurs when a project's cash inflows equal its cash outflows. Thus, the project's period-by-period operating cash flow is zero. The formula for the cash break-even point (Q_{cash}^*) is as follows:

$$Q_{\text{cash}}^* = \frac{FC}{p - vc}$$

where:

FC = fixed costs

vc = variable cost per unit

p = price per unit

For any project operating at cash break-even, net income (ignoring taxes) will equal depreciation expense. This stands to reason. Ignoring working capital for cash flow from operating activities, we know that operating cash flow (OCF) equals net income plus depreciation: $OCF = NI + \text{Dep}$. In the case of cash break-even, OCF is zero, so $NI = -\text{Dep}$.

Cash break-even tells us how much product must be sold so that the firm's overall operating cash flows are not reduced.

61. Cash Budget

A cash budget shows the cash flow that the firm anticipates in the upcoming period, given various scenarios. This budget goes beyond a simple summation to cash receipts and disbursements. Rather, it attempts to forecast the actual timing of the cash flows into and out of the business. The precision of the budget depends upon the characteristics of the organization, the degree of uncertainty about the business environment, and the ability of the planner to accurately forecast the future cash flows. The budget process is characterized by five steps:

1. Forecasting sales
2. Projecting all cash inflows, including forecasted receipts
3. Projecting all cash outflows
4. Interrelating the inflows and the outflows, subject to policy decisions of the firm's management
5. Determining the excess or shortage of cash during the period

Every cash budget must begin with a forecast of sales, which normally is supplied to the financial planner by the firm's marketing department. The primary source of cash inflow for many firms is not sales, but the collection of accounts receivable. In addition, the firm may raise cash from external sources through short-term or long-term financing or the sales of assets. These inflows also are part of the cash budget.

62. Cash Commodity

The actual physical commodity, as distinguished from a futures commodity. A commodity delivered at the time of sale is a cash commodity while a commodity to be delivered at a specific future date is a future commodity.

63. Cash Concentration Systems

A cash concentration system is designed to move funds from many small accounts into one or several large master accounts as efficiently as possible. A cash concentration network improves the financial manager's control of company cash by accumulating balances in one large account. The manager may be able to forecast total cash flows for the master account with a smaller percentage error than that associated with estimating cash balances of many small accounts. In addition, the manager can invest these funds at higher rates, since pooled funds can buy larger blocks of investment securities or money market instruments that are sold in large denominations. Finally, the cash concentration network can help reduce both excess balances in many small banks and expenses for transferring funds.

A concentration network uses DTCs (depository transfer check), wire transfers, and lockboxes to improve the efficiency of the firm's cash flows and investments. The type of system that a firm employs will depend upon the average dollar volume of its transactions, the number and sophistication of its banks, the timing and type of

information that it requires, and the current opportunity cost of float. For example, DTCs are preferable to wire transfers when transferring funds in small dollar amounts through a volume of transactions, since DTCs are much less expensive than wire transfers. However, a high volume of transactions involving disbursements that are known ahead of time (such as payroll) might induce the firm to use an Automated Clearing House (ACH) transfer. ACH transfers often handle high-volume transactions and regular (or batch) transactions, and they usually can make the funds available in one business day. Although the ACH cannot provide the same immediate availability as wire transfer, it is slightly less expensive than a DTC and may serve a useful purpose when handling certain types of payments. Cash concentration systems are improving firms' float management and information gathering. For example, ACH tapes now can be deposited on weekends to help reduce the firm's risk of overdrafts. Future cash concentration systems should continue to make strides in reducing excess balances, administrative costs, and transfer costs while providing the manager with more reliable information to help with investing cash and arranging appropriate lines of credit.

64. Cash Conversion Cycle

The cash conversion cycle is the net time interval between the actual cash outflow to pay accounts payable and the inflow of cash from the collection of accounts receivable.

The cash conversion cycle reflects the fact that some of the firm's inventory purchases are not immediately associated with cash outflows. Rather, the timeline shows that the firm buys inventories and then pays for them at some later time. Therefore, the cash conversion cycle is the distance on the timeline between payment for inventories and collection of accounts receivable:

$$\text{Cash conversion cycle} = \text{Operating cycle} - \text{Payable deferral period}$$

where

$$\begin{aligned} \text{Operating cycle} = & \text{Receivable collection period} \left(\frac{\text{Account receivable}}{\text{sales}/365 \text{ Days}} \right) \\ & + \text{Inventory conversion period} \left(\frac{\text{Inventory}}{\text{Cost of good sold}/365 \text{ days}} \right) \end{aligned}$$

A shorter cash conversion cycle makes a firm more liquid. This makes it an excellent tool by which to measure the overall liquidity of a firm. The cash conversion cycle helps the manager to model cash flow management decisions on a timeline to clearly show their effects. For example, if the firm introduces a new system to collect accounts receivable more quickly, the manager can compare the cash conversion cycles under the old and new systems to evaluate the effects of the new system. Other financial or operating decisions also can be incorporated into the cash conversion cycle framework to provide a method of analyzing their effects on the firm's cash flows.

The cash conversion cycle quantifies the time it takes for cash to flow out through the working capital accounts and back in to the cash accounts. Essentially, the cycle begins when the organization pays cash for an investment in current assets and ends when cash flows back to the organization as payment for its goods or services. The short-term financial planner's first task is to identify the firm's cash flow cycle. The next step is to focus on how to speed up inflows and slow down outflows in the most cost-effective fashion. If we use average inventory, average accounts receivable (AR), and average accounts payable (AP) to replace inventory, AR, and AP, the cash conversion cycle can be rewritten as:

$$\begin{aligned} \text{Cash conversion} &= \text{Average age of} + \text{Average age of} \\ &\quad - \text{Average age of} \\ &= \text{cycle inventory accounts receivable} \\ &\quad \text{accounts payable} \\ (\text{No. of days in planning period}) &\times \left[\frac{\text{Average inventory}}{\text{Cash operating expenditures}} \right. \\ &\quad \left. + \frac{\text{Average accounts recievable}}{\text{Sales}} - \frac{\text{Average accounts payable}}{\text{Cost of goods sold}} \right] \end{aligned}$$

where:

Average inventory = (Beginning inventory + Ending inventory)/2

Average accounts receivable = (Beginning AR + Ending AR)/2

Average accounts payable = (Beginning AP + Ending AP)/2

65. Cash conversion system

The cash conversion system consists of a treasury, a single-product make-to-stock inventory, and a receivables pool. It implements a perpetual flow cycle, where funds convert to product and back to funds. The system is stationary, and revenues and costs flow directly to the parent company. The parent company aims to maximize equilibrium (long-run) financial metrics in terms of net profit rate and rate of return. To this end, we model this system as a discrete-state continuous-time Markov process and compute its equilibrium state distribution using analytic and numerical methods. These are then used to derive statistics of the equilibrium cash conversion cycle and define equilibrium financial rate metrics and cumulative counterparts that incorporate the time value of money.

66. Cash Coverage Ratio

Cash coverage ratio is defined as the sum of EBIT and depreciation divided by interest. The numerator is often abbreviated as EBITDA.

67. Cash Cow

A company that pays out all earnings per share (EPS) to stockholders as dividends (Div). Hence, $EPS = Div$. The value of the share of stock becomes: $\frac{EPS}{r} = \frac{Div}{r}$, where r is the discount rate on the firm's stock. Cash cow project represents a strong market share and low market growth project.

68. Cash Cycle

In general, the time between cash disbursement and cash collection. In net working capital management, it can be thought of as the operating cycle less the accounts payable payment period. The cash cycle begins when cash is paid for materials and ends when cash is collected from receivables. [See **Cash conversion cycle**]

69. Cash Delivery

The provision of some futures contracts that requires not delivery of the underlying assets (as in agricultural futures) but settlement according to the cash value of the asset.

70. Cash Disbursement Systems

The primary purpose of a disbursement system is to minimize the net cost of delivering payments to a company's employees, suppliers, and stockholders. Such a system must consider several categories of cost: (1) Opportunity costs from investments not made or interest expenses for unnecessary borrowings. (2) Transfer costs associated with moving funds from one location to another. (3) Cost associated with lost discounts, or opportunity costs of late or early payments. (4) Costs associated with vendor/employee ill will. (5) Managerial costs of handling the disbursement system. (6) Costs of any unauthorized disbursements.

To reduce opportunity costs, a firm can design the system to increase disbursement float. [See **Disbursement float**] This can be accomplished, for example, by mailing checks from a remote disbursement location. The manager must balance the benefit of such a technique against the potential cost in strained relationships with vendors, though. Intentionally late payments or exaggerated mail float might create ill will among vendors and employees and cause the firm problems in the future.

71. Cash Discounts

The most obvious cost of a cash shortage comes from the inability to take advantage of suppliers' cash discounts by paying bills promptly. Most firms buy materials and supplies on terms of "2/10, net 30," which means that the buyer can deduct 2% from its bill if it pays within 10 days of receiving it, and that the payment in full is due within 30 days. Now, 2% may not sound like very much, but it allows you to use \$98 for 20 more days of credit (the difference between 10 days and

30 days). This is an effective rate of 2.04% (2/98). To realize the true cost of bypassing the discount, convert the percentage to an annual rate: 20 days is about one-eighteenth of a year, so the true rate is 2.04% times 18.25, or 37%. In other words, by paying its bills 20 days after the discount date, the company is in effect borrowing money at an annual interest rate of 37%.

The annualized cost of foregoing a discount can be found by the following general formula:

$$\begin{aligned} & \text{Annual cost of foregoing a discount} \\ &= \frac{\text{Percentage cash discount}}{100\% - \text{Percentage cash discount}} \\ & \quad \times \frac{365 \text{ days}}{\text{Date for net payment} - \text{Date of discount payment}} \end{aligned}$$

It may be argued that when a cash-poor company pays its bills late to stretch out its funds, very probably it will not pay even after 30 days. It will, in fact, pay as late as possible. If we assume that the firm pays suppliers' bills 60 days after receipt rather than 30 days, it exchanges the discount of 2% for 50 days of additional credit. This reduces the cost of foregoing discounts from 37% to about 15%. However, such a policy cannot be maintained indefinitely. It greatly harms the company's relations with its suppliers and possibly with the financial community as well. Additionally, firms paying later may face interest charges imposed by their suppliers. Any such practice should certainly be reserved for real emergencies.

72. Cash Equivalents

Short-term money-market securities. In general, the first item of current asset in a balance sheet is "cash or cash equivalents."

73. Cash Flows

Cash flows deal with the actual transfers of cash into or from the firm. Cash generated by the firm and paid to creditors and shareholders. It can be classified as (1) cash flow from operations, (2) cash flow from changes in fixed assets, and (3) cash flow from changes in net working capital.

74. Cash Flow After Interest and Taxes

Net income plus depreciation. It is also called net cash inflow in the capital budgeting decision.

75. Cash Flow from Operations

A firm's net cash flow from normal business operating activities used to assess the firm's ability to service existing and new debt and other fixed payment obligations.

76. Cash Flow Mapping

A procedure in which the cash flows of a given claim are assigned-or mapped-to a set of benchmark claims.

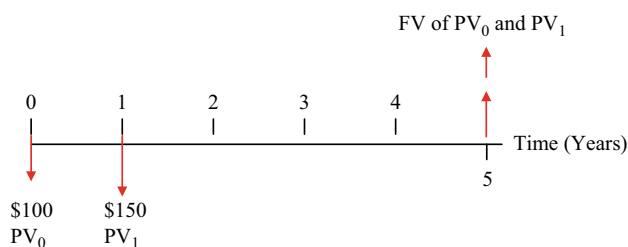
77. Cash Flow Matching

A form of immunization, matching cash flows from a bond portfolio with an obligation. [See **Dedication strategy**]

78. Cash Flow Timeline

A cash flow timeline can be a useful tool for visualizing and identifying cash flows over time. A cash flow timeline is a horizontal line with up arrows that represent cash inflows (i.e., cash to be received by the decision maker) and down arrows to indicate cash expenses or outflows. The down arrow at Time 0 represents an investment today; the up arrow n periods in the future represents $\$FV$, the future value (or compounded value) of the investment. For example, today \$100 is invested in a 5-year CD that advertises a 10% annual interest rate. One year later, an additional \$150 will be invested in a 4-year CD that pays 9%. How much money will be available when both CDs mature? Note that both CDs mature on the same date.

The cash flow timeline looks like this:



79. Cash Letter

Transit letter on tape that lists items submitted between banks for collection.

80. Cash-Market

The spot market for the immediate exchange of goods and services for immediate payment.

81. Cash Offer

Selling shares of seasoned equity to the public is called a cash offer. Usually, an investment bank is involved in the sale in one of three ways. A firm can issue

seasoned equity using a firm commitment underwriting, by selling all or part of a previously shelf-registered issue, or through a competitive bidding process.

82. Cash Offering

A method used by the US Treasury to sell securities in which buyers must pay cash to acquire new securities.

83. Cash Settlement

A procedure where settlement entails a cash payment from one party to the other, instead of delivery of an asset.

84. Cash Transaction

A transaction where exchange is immediate, as contrasted to a forward contract, which calls for future delivery of an asset at an agreed-upon price. It is a contrast to a forward contract. For example, if had the book been on the bookstore's shelf, you purchase of it would constitute a cash transaction.

85. Cashier's Check

A bank check that is drawn on the bank issuing the check and signed by a bank officer.

86. Cashout

Refers to situation where a firm runs out of cash and cannot readily sell marketable securities. It may need to borrow or default on an obligation. Therefore, cash management is very important for company to avoid the situation. [See **Baumol's economic order quantity model**]

87. Cash/Bond Selection

Asset allocation in which the choice is between short-term cash equivalents and longer-term bonds.

88. Cash-and-Carry

The simultaneous spot purchase and forward sale of an asset or commodity.

89. Cash-or-Nothing Call

An option that pays a fixed amount of cash if the asset price exceeds the strike price at expiration. If the asset price is equal or smaller than the strike price, then the call holder gets nothing.

90. Cat Bond

Bond where the interest and, possibly, the principal paid are reduced if a particular category of “catastrophic” insurance claims exceed a certain amount.

91. Cash-to-Cash Asset Cycle

The time it takes to accumulate cash, purchase inventory, produce a finished good, sell it, and collect on the sale.

92. Cash-to-Cash Liability Cycle

The length of time to obtain interest-free financing from suppliers in the form of accounts payable and accrued expenses.

93. Cash-to-Cash Working Capital Cycle

The timing difference between the cash-to-cash asset cycle and the cash-to-cash liability cycle.

94. CD Basis

CD basis is a method that results in a higher effective yield than the bond-equivalent basis. The math works like this: using daily compounding, the effective yield would be determined by dividing the annual rate by 360, and then compounding for 365 days. [See also **Bond-equivalent basis**]

95. Central Bank

The main bank in a country responsible for issuing currency and setting and managing monetary policy.

96. Central Limit Theorem

One of the most important results in statistics, which states that the sum of independent and identically distributed random variables has a limiting distribution that is normal.

97. CEO Compensation

Executive compensation is composed of both the financial compensation (executive pay) and other non-financial benefits received by an executive from their employing firm in return for their service. It is typically a mixture of fixed salary, variable performance-based bonuses (cash, shares, or call options on the company stock) and benefits and other perquisites all ideally configured to take

into account government regulations, tax law, the desires of the organization and the executive.

98. CEO Talent

CEO talent includes professional knowledge, leadership, innovation potential, etc.

99. Certainty Equivalent

The certain return providing the same utility as the risky return of a risky portfolio in terms of certainty equivalent coefficient (α). In other words, the intercept of an indifference curve which represents the certain return is called the certainty equivalent of the portfolios on that curve and in fact is the utility value of that curve. For example, the certainty equivalent method for capital budgeting under uncertainty has used a certainty equivalent coefficient to convert the risky net cash inflow into risk-free net cash inflow in terms of certainty equivalent coefficient (α). α is:

$$\alpha = \frac{\text{certain return}}{\text{risky return}}$$

where the value of α ranges from 0 to 1.

100. Certificates of Deposit

Short-term loans to commercial banks. There are active markets in CDs of 3-, 6-, 9-, and 12-month maturities.

101. Certification Effect

As with any other firm commitment offer, the investment bank carries the risk of price fluctuations after the primary market transaction. As with an IPO, this should increase investors' confidence. As an outside third party, the managing investment bank has examined the issuer and found the firm worth. The bank "puts its money where its mouth is" by giving a firm commitment price and underwriting the issue. This certification effect conveys information to the marketplace that the issue is fairly priced. The investment bank is staking its reputation and profits on the attractiveness of the issuer. Investment banking firms with the highest reputations (e.g., Goldman Sachs, Merrill Lynch, and Morgan Stanley) provide the strongest certification effects with respect to security sales. The certification effect provides a signal to the financial markets regarding the quality of the issuer.

102. Certified Check

A check guaranteed by a bank where funds are immediately withdrawn.

103. Certified Financial Planner (CFP)

A designation earned by individuals who have passed the examination sponsored by the Certified Financial Planner Board. Such individuals have studied banking, investment, insurance, estate planning, and tax planning to assist in managing client financial needs.

104. CES Production Function

The CES production function is a neoclassical production function that displays constant elasticity of substitution. In other words, the production technology has a constant percentage change in factor (e.g., labor and capital) proportions due to a percentage change in marginal rate of technical substitution.

105. Change in Net Working Capital

Difference between net working capital from one period to another. For example, the change in net working capital in 20×4 is the difference between the net working capital in 20×4 and 20×3 . The change in net working capital is usually positive in growing firm.

106. Changes in Fixed Assets

Component of cash flow that equals sales of fixed assets minus the acquisition of fixed assets. For example, when US Composite sold its power systems subsidiary in 20×4 it generated \$25 in cash flow.

107. Chapter 12

Chapter 12 of the Federal Bankruptcy Reform Act of 1978 tries to allow for a planned restructuring of the corporation while providing for payments to the creditors. Chapter 12 proceedings begin when a petition is filed by the corporation or by three or more creditors. A federal judge either approves or disapproves the petition for protection under Chapter 12. During the petition period, the judge protects the managers and shareholders from the creditors and tries to negotiate a rescue plan between the shareholders and creditors. During this time, the corporation continues to do business.

Once in Chapter 12, the firm's management has 120 days to submit a reorganization plan, which usually includes debt rescheduling and the transfer of equity rights. Anyone has the right to submit such a plan, but only very rarely does anyone but management submit a reorganization plan. The plan must secure the agreement of two-thirds of the shareholders and two-thirds of each class of creditors; for example, senior creditors whose debt is secured and junior creditors whose debt is unsecured are considered separate classes.

After the plan is approved, the judge confirms it. At this point, any payments, property sales, or securities issues or transfers of equity positions take place under the supervision of the court.

Some critics argue that Chapter 12 is flawed and needs reform because it favors shareholders over creditors and junior creditors. They claim it is unfair that shareholders and junior creditors can vote to approve the reorganization plan as equals with the senior creditors.

Also, time works against creditors in Chapter 12. Upon approval of a reorganization plan by the court, interest payments to creditors stop and legal fees begin to erode the remaining value of the firm. Often, senior creditors settle for less than their full debts simply to save time. Shareholders, on the other hand, wish to draw out the reorganization period as long as possible hoping for a turnaround; they have little or nothing left to lose.

In general, this delay is bad for the company. If a firm's managers know they can default on debts and still keep their jobs, they may tend to abuse creditors. This could cause shareholders to require larger returns on their capital and creditors to be less willing to risk their funds.

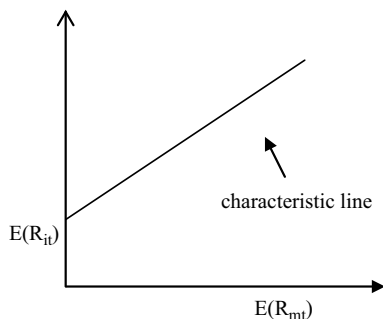
Critics have presented two basic ideas for reforming Chapter 12: (1) increase the bureaucracy, and (2) allow the market to decide. The first proposes setting time deadlines after which independent arbitrators (more bureaucracy) decide a firm's fate. This would put bankruptcy more firmly in the hands of bureaucrats. The second reform proposal involves creating opportunities for creditors and owners to sell their positions to each other or third parties at prices determined competitively in the market. This market-based solution would encourage whoever ends up with equity control to make the firm as valuable as possible.

108. Chapter 8

Chapter 8 of the Bankruptcy Reform Act of 1978 covers the **liquidation** of a firm. [See also **Liquidation**]

109. Characteristic Line

The line relating the expected return on a security, $E(R_{it})$ to different returns on the market, $E(R_{mt})$. This is a straight line plotting in the dimension with X-axis as % in return on market, Y-axis as % return on security. The slope of characteristic line is the beta. [See also **Market model**]



110. Charge-Off

The act of writing off a loan to its present value in recognition that the asset has decreased in value.

111. Charter

A legal document that authorizes a bank to conduct business.

112. Chartered Financial Analyst (CFA)

A designation earned by individuals who have passed a three-part examination sponsored by the Institute of Chartered Financial Analysts. Topics include economics, finance, security analysis, and financial accounting to assist in security analysis and portfolio management.

113. Chartists

Some investors, called chartists or **technicians**, examine graphs of past price movements, number of shares bought and sold, and other figures to try to predict future price movements.

114. Cheapest to Deliver

When a futures contract permits the seller to select the precise asset or commodity to deliver to the buyer, the cheapest to deliver is the asset that is most profitable for the shorter to deliver.

115. Check Kiting

The process of writing checks against uncollected deposits while checks are in the process of collection, thereby using funds (float) not actually available.

116. Chief Financial Officer (CFO)

In all but the smallest of firms, a top manager with the title chief financial officer (CFO) or vice president of finance usually reports to the president. Managers of two areas usually report to the CFO: the firm's **treasurer** and its **controller**. [See also **Treasurer** and **Controller**]

The CFO serves as the heir apparent for the CEO in many organizations.

117. Chinese Wall

The imaginary barrier that ensures a trust department will manage trust assets for the benefit of the trust beneficiaries, not for other departments in the bank.

118. Chooser Option

An option where the holder has the right to choose whether it is a call or a put at some point during its life. This prespecified date is normally called the choice date. The chooser options can either a standard simple chooser option or complex chooser option. The former refers to both call and put are specified with same price and maturity time. The latter refers to call price and maturity date are different from those of put option.

119. Class of Options

A class of options refers to all call and put contracts on the same underlying asset. For example, all AT&T call and put options at various exercise prices and expiration months form one class. [See **Option class**]

120. Classic Hedge Strategy

The implicit assumption of the classic hedge ratio equal to one is that the prices of the spot commodity (in this case, the stock portfolio) and the futures contract will remain perfectly correlated over the entire hedge period. Then if the stock market does turn down as expected, as losses in the portfolio due to price declines in its composite stocks will be exactly offset by the gain on the futures position. Conversely, if stock prices rise, the portfolio's gain will be offset by equal losses on the future position. Such a strategy implies that the objective of the classic hedge is risk minimization or elimination.

121. Classical Theory of Interest Rates

An explanation of the level of and changes in interest rates that relies upon the interaction of the supply of savings and the demand for investment capital.

122. Classical Time-Series Decomposition

Classical time series decomposition separates a time series into five components: mean, long-range trend, seasonality, cycle, and randomness.

123. Clean Price of Bond

The quoted price of a bond. The cash price paid for the bond (or dirty price) is calculated by adding the accrued interest to the clean price.

124. Clearing

The exchanging of checks and balancing of accounts between banks.

125. Clearing Margin

A margin posted by a member of a clearinghouse.

126. Clearinghouse

The third party of every futures contract, which guarantees that every futures contract will be carried out even if one of the parties defaults. The clearinghouse also facilitates trading of futures contracts before they are due for delivery.

127. Clearinghouse Automated Payment System (CHAPS)

The Clearinghouse Automated Payment System (CHAPS) is a large-value, electronic credit transfer system that provides same-day funds transfers for British pound payments. Located in London, the clearing network is similar to Clearinghouse Interbank Payment System (CHIPS). The 14 CHAPS members operate the system and settle balances at the end of each day through the Bank of England. CHAPS offers transfer services to other banks and customers through its 14 members.

Transfers through CHAPS are considered final; they are guaranteed, irrevocable, and unconditional. Remember, however, that this is a private system, and thus the transfers are guaranteed by the members and not by the Bank of England.

128. Clearinghouse Association

A voluntary association of banks formed to assist the daily exchange of checks among member institutions.

129. Clearinghouse Funds

Money transferred by writing a check and presenting it for collection.

130. Clearinghouse Interbank Payment System (CHIPS)

The Clearinghouse Interbank Payment System (CHIPS) is a private payment-clearing system located in New York City and operated by the New York Clearinghouse Association. The network specialized in international payments denominated in US dollars. It is estimated that CHIPS transfers 90% of all international, interbank dollar transactions.

CHIPS tabulates transaction data for all member banks at the end of each day. The system also permits members to review payments online, that is, those in storage awaiting approval. This gives member banks better information about available funds on which to base their credit decisions. A bank officer may be more willing to grant credit knowing that CHIPS will credit a certain account later in the day.

Although the system is highly technical, CHIP's membership, operating procedures, and efficiency have important economic implications. For one, the role of

the US dollar as a world transaction currency is influenced by the relative operating efficiency and safety of its payment mechanism. Further, depending upon CHIP's handling of failures to settle accounts, such an occurrence can be either an isolated event, or the first link in a chain reaction leading to a worldwide liquidity crisis. Last, but not least, the specific roles of individual banks in the dollar clearing system have important implications for worldwide correspondent banking relationships and, therefore, market share and profits.

131. Clientele Effect

Both federal and state governments tax dividend income at ordinary income tax rates. Any differences between capital gains and income tax rates will lead some investors to prefer one or the other for tax reasons. Another influence on dividend policy will be the composition of the firm's shareholders, commonly called its *clientele*.

Miller and Modigliani introduced the clientele effect as an imperfection of the market that affects dividend policy. M&M observed that each corporation tends to attract a specific type of clientele that favors the firm's established payout ratio. For example, investors in higher tax brackets tend to hold stocks with lower dividend payouts and higher capital gains yields. This way, they can avoid personal taxes on dividend income. On the other hand, retirees, because of their lower tax brackets, tend to invest in companies with larger yields.

Because a firm tends to attract a certain type of investor, management may be reluctant to change its dividend policy. If shareholders have to change their portfolios due to changes in payout ratios, this shift may cause shareholders to incur unwanted transaction costs. In this way, the tax differential favoring capital gains is a systematic imperfection of the market that produces a clientele effect.

132. Closed-End Country Funds

Closed-end country funds are funds that issue a fixed number of shares and invest the proceeds largely in equity securities from a particular (local) country. It is generally called country funds.

133. Closed-End (Mutual) Fund

A fund whose shares are traded through brokers at market prices; the fund will not redeem shares at their net asset value. The market price of the fund can differ from the net asset value. For example, a country fund such as Korean fund is a closed-end fund. In addition, REIT is also a closed end fund. [See **REIT**]

134. CMO

[See **Collateralized mortgage obligation**]

135. Cobb-Douglas Production Function

In economics and econometrics, the Cobb–Douglas production function is a particular functional form of the production function, widely used to represent the technological relationship between the amounts of two or more inputs (particularly physical capital and labor) and the amount of output that can be produced by those inputs. The elasticity of substitution of the Cobb-Douglas production function is equal to 1.

136. Coefficient of Determination

[See **R-squared** (R^2)]

137. Coefficient of Variation

One problem with using the standard deviation as a measure of risk is that we cannot get an intuitive feel for risk by looking at the standard deviation alone. Firm A's profits may have a higher standard deviation than firm B, but because firm A's mean return is much higher, firm A actually may have lower risk. The coefficient of variation allows us to make comparisons because it controls for the size of the average. The coefficient of variation (CV) measures risk-per-unit of return. The coefficient of variation is computed as the standard deviation divided by the mean:

$$CV = \sigma / \bar{X}$$

138. Coincident Indicators

Economic indicators can be classified as leading, coincident, and lag indicators. A coincident indicator is a metric that shows the current state of economic activity within a particular area. Coincident indicators are important because they show economists and policymakers the current state of the economy. Coincident indicators include employment, real earnings, average weekly hours worked in manufacturing, and the unemployment rate.

139. Collar

Use of options to place a cap and floor on a firm's borrowing costs. One way to do this is to sell a floor and use the premiums on the floor to pay the premium on the purchase of the cap. [See **Interest rate collar**]

140. Collar

A collar option strategy, also referred to as a hedge wrapper or simply collar, is an option strategy employed to reduce both positive and negative returns of an underlying asset. It limits the return of the portfolio to a specified range and can hedge a

position against potential volatility of the underlying asset. A collar position is created through the usage of a protective put and covered call option. More specifically, it is created by holding an underlying stock, buying an out of the money put option, and selling an out of the money call option.

141. Collar Width

The difference between the strike prices of the two options in a collar.

142. Collateral

One way in which a bank can limit its exposure to risk is by requiring the borrower to pledge some valuable assets as collateral – that is, security for the loan. For example, a company that produces buildings, locomotive, large generating plants, or other major pieces of equipment may pledge these high-value items as security.

143. Collateralized Bonds

Collateralized bonds pledge securities to protect bondholders against loss in case of default. An example of collateralized bonds are collateralized mortgage obligations (CMOs) sold by firms and agencies involved in the housing market; the CMO is backed by a pool of mortgages. Other examples of collateralized bonds include bonds backed by credit card receivables and bonds backed by car loans. The issuer pays interest and principal on such a collateralized bond over time as homeowners, credit card users, and car buyers pay off their own loans.

144. Collateralized Debt Obligation

A way of packaging credit risk. Several classes of securities are created from a portfolio of bonds and there are rules for determining how defaults are allocated to classes.

145. Collateralized Mortgage Obligation (CMO)

A security backed by a pool of mortgage that is structured to fall within an estimated maturity range (tranche), based on the timing of allocated interest and principal payments on the underlying mortgages. A CMO services as a way to mitigate or reduces prepayment risk of a real estate loan. [See **OAS**]

146. Collateral Trust Bond

A bond secured by a pledge of common stock held by the corporation.

147. Collect-on-Delivery Option

An option where the premium is paid only when the option is exercised.

148. Collected Balances

Ledger balances minus float. [See also **Float**]

149. Collection Float

An increase in book cash with no immediate change in bank cash, generated by checks deposited by the firm that have not cleared. [See **Float**]

150. Collection Policy

Procedures and policy followed by a firm in attempting to collect accounts receivable. It is one of the components of credit policy.

151. Combination

A position involving both calls and puts on the same underlying asset.

152. Combined Investment Strategy

Combined investment strategy involves the combination of bonds, stock, futures, options, etc. to reduce risk and increase returns.

153. Combined Leverage

Operating leverage and financial leverage combine to magnify a given percentage change in sales to a potentially much greater percentage change in earnings. Together, operating and financial leverage produce an effect called combined leverage. [See **Degree of combined leverage**]

154. Commercial Bank

A business corporation that accepts deposits, makes loans, and sells other financial services, especially to other business firms, but also to households and governments.

155. Commercial Draft

A commercial draft resembles a promissory note, but it works somewhat differently. First, the seller draws a draft ordering payment by the customer and sends this draft to the customer's bank along with any shipping documents. This

commercial draft is called a *sight draft* if immediate payment is required; otherwise, it is a *time draft*, on which the customer's signature and the word *accepted* must be added. In either case, the advantage of this trade-credit instrument is that the seller obtains the buyer's formal commitment to pay before goods are delivered. This commitment is the money that the seller receives ahead of time, or the trade acceptance the buyer signs, which the bank then returns to the seller. In sum, commercial draft is a demand for payment.

156. Commercial Loan

A theory suggesting that banks make only short-term, self-liquidating loans that match the maturity of bank deposits.

157. Commercial Mortgage

A debt obligation whose purpose is to finance the construction of office buildings, shopping centers, and other business structures.

158. Commercial Paper

Large companies have a very attractive source of short-term funds open to them: they can sell commercial paper, unsecured promissory notes that trade in the organized money market through a number of recognized dealers. The buyers of the paper are primarily commercial banks looking for safe investments that yield higher returns than US Treasury securities. Other buyers include corporations, pension funds, insurance companies, and others that have temporary surplus funds they wish to put to work safely.

Commercial paper is sold in two ways: (1) the issuer may sell the paper directly to the buyer, or (2) the issuer may sell the paper through a dealer firm. Firms prefer to sell directly to save the dealer's fee of approximately one-eighth of a percentage point (12.5 basis points). One hundred basis points equal 1%. Commercial paper is sold on a discount basis. Almost half of commercial paper is issued directly, with most of the direct paper being issued by finance companies. Approximately 75% of all paper (both direct and dealer issues) comes from financial companies, including commercial, savings, and mortgage banking firms, finance leasing, insurance underwriting, and other investment activities. The balance of outstanding paper is issued by nonfinancial firms, such as utilities and industrial manufacturers. This paper ordinarily is issued by a dealer.

Beside its relative low cost, commercial paper offers three advantages. First, selling the notes is a fairly simple and informal process, certainly simpler than negotiating a bank loan. While it is not as easy as using trade credit, commercial paper is the simplest of all forms of negotiated credit. Second, the ability to sell unsecured promissory notes gives the issuer a degree of prestige. This, in turn, makes

it even easier to sell later issues as the company builds a name for itself in the money market. Third, a commercial paper issue may exceed the legal lending limit of most commercial banks, preventing the need to combine banks to assemble a financing package.

At first sight, commercial paper may seem to be the obvious choice because of its lower cost, but reliance on commercial paper may be a high-risk policy. A company that finances all of its short-term needs through the sale of notes does not build up a good borrowing relationship with a bank. If economic conditions change and the money market becomes tight, such a company may well find itself in difficulties. The banks will give priority to their regular customers; they may not even have enough loanable funds to meet all the needs of their regular borrowers. The company that has relied on the money market when money was easy will have to continue to rely on it when funds are scarce, and the differential between the interest rates of the two sources is likely to shrink dramatically in such circumstances.

In sum, commercial paper is a short-term, unsecured promissory notes issued by corporation with a high credit standing. Their maturity ranges up to 270 days.

159. Commission Broker

A broker on the floor of the exchange who executes orders for other members.

160. Commitment

A legally binding obligation (subject usually both to conditions precedent and to continuing conditions) to make available loans or other financial accommodation for a specified period; this includes revolving facilities. Even during publicly known credit distress, a commitment can be legally binding if drawn down before it is formally withdrawn for cause.

161. Commitment Fee

Fee charged for making a line of credit available to a borrower.

162. Committed Line of Credit

With the committed line of credit, the borrower pays an up-front fee which then obliges the bank to lend the firm money under the terms of the line of credit. [See also **Revolving credit agreement**]

163. Commodity Futures Trading Commission

A body that regulates trading in futures contracts in the United States.

164. Commodity Diversifier

An investment in commodities could be timely, given the current relationship to historical prices. A tactical approach that mitigates risk makes sense in light of market uncertainties and the potential for dramatic price swings.

165. Commodity-Indexed Bonds

Several firms have issued commodity-indexed bonds with exposure to prices of commodities such as oil, gold, or silver. In a way, this technique closely resembles the tactic by which a US firm hedges its overseas risk by issuing bonds denominated in a foreign currency. An oil-drilling firm's cash flows are sensitive to the price of oil, as are the cash flows of gold- or silver-mining firms to the prices of those commodities. By making coupon interest and/or principal amounts vary along with the commodity price; these firms can reduce their risk of bankruptcy. Falling commodity prices reduce such a firm's cash flows, so its debt service requirements decline, as well.

166. Commodity Spread

Offsetting long and short positions in closely related commodities. [See also **Crack spread** and **Crush spread**]

167. Commodity Swap

A swap where cash flows depend on the price of a commodity.

168. Common-Base-Year Financial Statements

To see how the ledger items change over time, we can choose a base year balance sheet or income statement and then express each item relative to the base year. Such statements are referred to as common-base-year statements.

169. Common-Size Financial Statements

Common-size financial statements include common-size balance sheets and common-size income statements. A common-size balance sheet expresses all balance sheet accounts as percentages of total assets. A common-size income statement expresses all income statement items as percentages of gross sales.

Common-size statements give an analyst insight into the structure of a firm's financial statements at a particular point in time or during a particular period. They also indicate the percentage of sales consumed by production costs or interest expenses, the proportion of assets that are liquid, or the proportion of liabilities that are short-term obligations.

170. Common Stock

Equity claims held by the “residual owners” of the firm, who are the last to receive any distribution of earnings or assets. It is usually applied to stock that has no special preference either in dividends or in bankruptcy. Owners of common stock in a corporation are referred to as shareholders or stockholders. They receive stock certificates for the shares they own. Owners of common stock are responsible for the election of the Board of Directors, appointment of Senior Officers, the selection of an auditor for the corporate financial statements, dividend policy, and other matters of corporate governance.

171. Common Stock Equivalents

Because of the possible dilution in earnings per share (EPS) represented by securities that have the potential to become new shares of common stock, the EPS calculation must account for common stock equivalents (CSEs). CSEs are securities that are not common stock but are equivalent to common stock because they are likely to be converted into common stock in the future. Convertible debt, convertible preferred stock, stock rights, stock options, and stock warrants all are securities that can create new common shares and thus dilute (or reduce) the firm’s earnings per share.

172. Community Reinvestment Act

A federal law passed in 1977 that requires depository institutions to designate the market areas they will serve and to provide those services without discrimination to all neighborhoods and regions within their designated market areas.

173. Comparative Static Analysis for Option Pricing Model

It is a sensitive analysis of option pricing model by taking partial derivative with respect to current stock price per share, exercise price and contract period, standard deviation of rates of return, and risk-free rate. [See **Delta, Theta, Vega, Rho, and Gamma**]

174. Comparison Universe

The collection of money managers of similar investment style used for assessing relative performance of a portfolio manager.

175. Compensating Balances

A company’s cash needs fall into three categories: (1) cash for day-to-day transactions, (2) reserve cash to meet contingencies, and (3) cash for

compensating balance requirements. A compensating balance exists when a firm must keep minimum cash balance in a noninterest bearing account at a bank as a condition or a loan or bank service agreement. To determine the appropriate minimum cash balance, a financial manager simply adds together the three segments just estimated. If the cash budget projects a balance significantly higher than the minimum balance, the organization can invest the excess cash in marketable securities. On the other hand, if the cash balance falls below the desired level, the organization can plan to sell marketable securities or to borrow short-term funds. To complete the transition from a cash flow budget to a cash flow plan, the manager must adjust the cash balance to meet the minimum cash balance.

In other words, compensating balance is a deposit that the firm keeps with the bank in a low-interest or non-interest-bearing account to compensate banks for bank loans or service.

176. Competitive Banking Equality Act

A 1987 federal law that requires depository institutions to speed up giving customers credit for deposits made in their checking accounts, placed a moratorium on the offering of new banking services, and gave federal insurance agencies additional tools to deal with failing depository institutions.

177. Competitive Bidding Issue

A competitive bidding issue occurs when a firm announces the size and terms of a proposed security sale and then solicits bids from investment banks to underwrite the issue. Once it accepts a bid, the firm proceeds with the sale just as for a firm commitment underwriting. The competition among banks may reduce the costs of floating the issue. Competitive bid underwriting involves no positive certification effect, as a bank must commit to a price before it can adequately perform its due diligence review and investigate the issuer. Unfortunately, few firms other than US utility companies and French public companies sell seasoned equity by competitive bidding (both of these classes of firms are required by law to seek bids to float security issues).

178. Competitive Offer

Method of selecting an investment banker for a new issue by offering the securities to the underwriter bidding highest. In a competitive bid process, the issuer, usually with the assistance of a financial advisor, structures the bond issue and publishes a notice of sale requesting bids from underwriters. After the bids are received, the bonds are awarded to the underwriting syndicate that submitted the best bid (i.e., the lowest true interest cost to the issuer).

179. Complete Portfolio

The entire portfolio, including risky and risk-free assets.

180. Complex Capital Structure

A corporation that has warrants, convertibles, or options outstanding is said to have a complex capital structure. The complexity comes from the difficulty of measuring the number of shares outstanding. This is a function of a known amount of common shares currently outstanding plus an estimate of the number of shares that may be issued to satisfy the holders of warrants, convertibles, and options should they decide to exercise their rights and receive new common shares.

181. Component Analysis

It is one of the two major approaches to time-series analysis. Component analysis regards the time series as being composed of several influences or components that are generally taken to be trend-cycle, seasonal, and random movement. The seasonal and trend movements are modeled in a deterministic manner. This kind of analysis is easier than the sample-function analysis. [See **Sample-function analysis**]

182. Composition

Composition is a way a firm can adjust its capital sources. This method involves recomposing the debt of the firm in such a way that the creditors receive partial payment for their claims, say, 60 cents for each dollar. Creditors may find it more expedient to follow this route than to take the troubled firm to court to seek full satisfaction. In court, they would run the risk of receiving less than they would through composition. Moreover, court appearances require various legal costs, which may more than offset the possible gains achieved by going to court.

In sum, composition is a voluntary arrangement to restructure a firm's debt, under which payment is reduced.

183. Composite-Based Beta Forecasting

Lee et al. (1986, pp. 51–68) proposed composite-based beta forecasting method. The composite-based beta forecasting is the weighted average of the accounting based and market-based beta forecasting. [See **Accounting-based** and **Market-based beta forecasting**]

184. Compound Interest

Interest that is earned both on the initial principal and on interest earned on the initial principal in previous periods. The interest earned in one period becomes in

effect part of the principal in a following period. The longer-lasting the loan, the more important interest on interest becomes.

$$\text{Future value of an investment : } FV = C_0(1 + r)^T$$

where C_0 is the cash to be invested at date 0; r is the interest rate; and T is the number of periods over which the cash is invested.

185. Compound Option

An option that has an option on the underlying asset. There are four main types of compound option: a call on a call, a call on a put, and a put on a put. Compound options have two strike prices and two exercise dates.

186. Compounding

Compounding involves finding the future value of money invested today. In other words, compounding allows us to determine how money will grow over time. The future value of a cash flow (PV) invested today (PV_0) for n periods at r percent interest per period is given by:

$$FV_n = PV_0(1 + r)^n$$

Compounding can either be discrete or continuous. [See also **Continuous compounding**]

187. Compounding Frequency

This defines how an interest rate is measured. Frequency can be daily, weekly, monthly, quarterly, annually, or continuously. [See **Continuous compounding**]

188. Compounding Swap

This kind of option is a variation on the plain vanilla swap. Swap where interest compounds instead of being paid. In other words, the interest is compounded forward until the end of the life of the swap.

189. Compound Value

Value of a sum after investing it over one or more periods. Also called future value.

[See **Future value**]

190. Computational Finance

Computational finance is a technology using computers, mathematics, and statistics to analyze finance strategy and investment decisions.

191. Concave Function

A concave function is one shaped like the cross section of an upside-down bowl. For example, a function used to describe the relationship between yield to maturity and years to maturity is generally a concave function. This is because the yield to maturity for a long-term bond is higher than that of a short-term bond.

192. Concave Utility Function

In mathematics, a concave function is the negative of a convex function. A concave function is also synonymously called concave downwards, concave down, convex upwards, convex cap, or upper convex. A real-valued function on an interval or a convex set in a vector space can be concave. The sum of two concave functions is itself concave and so is the pointwise minimum of two concave functions, as the set of concave functions on a given domain forms a semi field.

193. Concentration Banking

One way to speed up the collection of payments from customers is through concentration banking. In such a system, customers in a particular location make their payments to a local branch office rather than to company headquarters. The branch office then deposits the checks into a local bank account. The firm can then transfer surplus funds periodically to one of the firm's principal banks, called *concentration banks*.

This system reduces mail, processing, and collection float. However, concentration banking involves some additional costs, such as higher administrative costs, compensating balances required by the local bank for its services, and the cost associated with transferring the funds from the local bank to the concentration bank.

In sum, concentration banking is the use of geographically dispersed collection centers to speed up the collection of accounts receivable.

194. Concentration Risk

Portfolio risk resulting from increased exposure to one obligor or groups of correlated (e.g., by industry or location) obligors.

195. Conditional Autoregressive Range (CARR) Model

The CARR model provides an interesting alternative to the traditional volatility models (e.g., GARCH and stochastic volatility). We derive conditions for the existence of the first moment, stationarity, geometric ergodicity and beta-mixing property with exponential decay for the multivariate CARR. Empirical applications, based on market indices, show that the CARR model with generalized beta type II distribution provides the most efficient modeling of volatility for all data. This model is frequently used to predict the market rate of return or individual stock rates of return.

196. Conditional Sales Contract

An arrangement whereby the firm retains legal ownership of the goods until the customer has completed payment. A firm uses it as a credit instrument. Conditional sales contracts usually are paid off in installments and have interest costs built into them.

197. Conditional Value at Risk (C-VaR)

Expected loss during N days conditional on being the $(100 - X)\%$ tail of the distribution of profits/losses. The variable N is the time horizon, and $X\%$ is the confidence level.

198. Confidence Index

The confidence index is designed to measure how willing investors are to take a chance in the market. It is the ratio of high-grade bond yields to low-grade bond yields. This ratio is started below 1. When bond investors grow more confident about the economy, they shift their holdings from high-grade to lower-grade bonds, lowering their yield relative to high-grade bonds and increasing the confidence index. In other words, the confidence ratio moves close to 1.

Confidence-index technicians believe that the confidence index leads the stock market by 2–11 months. An upturn in the confidence index is supposed to foretell of rising optimism and rising prices in the stock market. A fall in the confidence index represents the fact that low-grade bond yields are rising faster or falling more slowly than high-grade yields. This is supposed to reflect increasing risk aversion by institutional money managers who foresee an economic downturn and rising bankruptcies and defaults. Analysts who have examined the confidence index conclude that it conveys some information for security analysis.

199. Confirmation

Contract confirming verbal agreement between two parties to trade in the over-the-counter market.

200. Conflict Between Bondholders and Stockholders

These two groups may have interest in the corporation that conflict. Sources of conflict include dividends, dilution, distortion of investment, and underinvestment. Protective covenants work to resolve these conflicts. Stockholders and bondholders have different objective functions, and this can lead to agency problems, where stockholders can expropriate wealth from bondholders. And because the firm is interested in trying to maximize stockholders wealth, there can develop a conflict of interest between stockholders and bondholders. For instance, stockholders have an incentive to take riskier projects than bondholders do and to pay more out in dividends than bondholders would like them to. This conflict can lead to costly decisions by the firm, which lowers the total value of the firm.

201. Conglomerate Acquisition

Acquisition in which the acquired firm and the acquiring firm are not related, unlike a horizontal or a vertical acquisition. For example, the acquisition of a food products firm by a computer would be considered a conglomerate acquisition.

202. Conglomerate Combination

A conglomerate combination is a type of business combination that may involve firms that have little, if any, product market similarities. A firm that is called a *conglomerate*, however, generally is one that has engaged in several conglomerate combinations.

203. Consensus Forecast

A prediction of interest rates or economic conditions based on a variety of projections derived from several different forecasting methods.

204. Conservator

An individual or trust department appointed by a court to manage the property of an incapacitated individual.

205. Consol

Consols are bonds that never stop paying a coupon, have no final maturity date, and therefore never mature. Thus, a consol is perpetuity. British bonds are called consols. [See **Perpetuity**]

206. Consolidated Balance Sheet

A balance sheet showing the aggregate financial condition of a firm and its subsidiaries, netting out all intracompany transactions.

207. Consolidation

Assuming there are originally two firms, Firm A and Firm B. One possible business combination results in the formation of a new firm, Firm C, which has the assets of both Firm A and Firm B. This is called a consolidation.

208. Constant Dividend Growth Model

[See **Gordon model**]

209. Constant Elasticity-of-Variance of Option Pricing Model

This model allows the variance term to follow a diffusion process in which its elasticity is always constant. It will allow the variance of the rate of return to vary inversely with stock price. In 1989, Professor Schroder in his *Journal of Finance* (pp. 211–220) has shown this kind of call option model can be defined as

$$\begin{aligned}
 C = & \text{(current stock price per share)} \\
 & \times \text{(first cumulative density function of noncentral } \chi^2) \\
 & - \text{(present value of exercise price)} \\
 & \times \text{(second cumulative density function of noncentral } \chi^2)
 \end{aligned}$$

This kind of option pricing model can be reduced to Black-Scholes option pricing model. [See **Black-Scholes formula**]

210. Constant Growth Model

A form of the dividend discount model that assumes dividends will grow at a constant rate. [See also **Gordon model**]

211. Constant Maturity Swap (CM Swap)

A swap where a swap rate is exchanged for either a fixed rate or a floating rate on each payment date. For example, an agreement to exchange 6-month LIBOR rate for the 10-year swap rate every 6 months for next 6 years.

212. Constant Maturity Treasury Swap (CMT Swap)

A swap where yield on a Treasury bond is exchanged for either a fixed rate or a floating rate on each payment date. For example, agreement to exchange a LIBOR rate for Treasury rate (e.g., the 10-year treasury rate).

213. Consumer Bank

A bank that does not make commercial loans.

214. Consumer Cash Management Service

A financial service in which an individual can hold surplus cash funds in an interest-bearing payments account until investments in higher-yielding assets can be made and through which credit can be obtained.

215. Consumer Credit

Credit granted to consumers. Trade credit is credit granted to other firms.

216. Consumer Sentiment

Consumer sentiment is a statistical measurement and economic indicator of the overall health of the economy as determined by consumer opinion. Consumer sentiment considers an individual's feelings toward his or her current financial health, the health of the economy in the short-term and the prospects for longer-term economic growth.

217. Constructive Sale

A term in tax law describing the owner of an asset entering into an offsetting position that largely eliminates the risk of holding the asset.

218. Consumption Asset

An asset held for consumption rather than investment.

219. Consumption-Based Asset Pricing Model

The consumption capital asset pricing model (CCAPM) is an extension of the capital asset pricing model (CAPM) that uses a consumption beta instead of a market beta to explain expected return premiums over the risk-free rate. The beta component of both the CCAPM and CAPM formulas represents risk that cannot be diversified away. The consumption beta is based on the volatility of a given stock or portfolio. The CCAPM predicts that an asset's return premium is proportional to its consumption beta.

220. Contagion

Financial contagion is traditionally defined as a situation in which a shock that initially affects only a few financial institutions spreads, or spills over, to the rest of the financial system and the economy, subsequently infecting the financial systems and economies of other countries. Because of its clear importance to the global economy, this phenomenon is one of the most widely researched areas of finance and has implications for portfolio management, trading, hedging, and diversification strategies.

221. Contemporaneous Reserve Accounting

The method of determining the amount of legal reserves a bank or other depository institution must hold behind its deposits and other reservable liabilities in which the reserve computation and reserve maintenance periods overlap.

222. Contingent Claim

Claim whose value is directly dependent on, or is contingent on, the value of its underlying assets. For example, the debt and equity securities issued by a firm derive their value from the total value of the firm. When the value of the firm exceeds the amount promised to the debtholders, the shareholders obtain the residual of the firm's value over the amount promised the shareholders, and the debtholders obtain the amount promised. When the value of the firm is less than the amount promised the debtholders, the shareholders receive nothing and the debtholders get the value of the firm.

223. Contingent Immunization

A mixed passive-active investment strategy that immunizes a portfolio if necessary to guarantee a minimum acceptable return but otherwise allows active management.

224. Contingent Liabilities

Items, such as guarantees or related contracts, that may become liabilities if certain developments arise.

225. Contingent Pension Liability

Under ERISA, the firm is liable to the plan participants for up to 30% of the net worth of the firm.

226. Continuous Compounding

An extreme example of frequent compounding is continuous compounding. Continuous compounding has many financial applications. These range from determining future values on bank accounts that advertise continuous compounding to topics such as the Black-Scholes option pricing model used by most option traders.

We know that $FV_n = PV_0 \left(1 + \frac{r}{m}\right)^{mn}$, that is, the future value (FV) of an investment today is equal to the amount invested multiplied by a future value interest factor that reflects compounded interest. Note that r is the interest rate per period and m is the number of compounding periods each year, and n is the number of years in the investment horizon.

The future value interest factor $[(1 + r/m)^m]$ rises at a decreasing rate and asymptotically approaches the continuous compounding FVIF of e^r . The FVIF rises as m increases, as the effect of more frequent compounding is to raise the effective annual rate (EAR). Higher EARs result in larger future values.

In sum, continuous compounding implies that interest compounded continuously, every instant, rather than fixed intervals.

227. Continuous Discounting

The present value of interest factor (PVIF) $= [1/(1 + r/m)^m]$ decreases at a decreasing rate and asymptotically approaches the continuous compounding FVIF of e^{-r} . The PVIF falls as m increases, as the effect of more frequent compounding is to raise EAR. Higher discount rates result in lower present values. [Notation definitions see **Continuous compounding**]

228. Continuously Compounded Interest Rate

A way of quoting an interest rate such that if \$1 is invested at a continuously compounded rate of r , the payoff in 1 year is e^r .

229. Contracting Costs

[See **Transaction costs**]

230. Contract Amount

The number of units of the good or service to be delivered.

231. Contract Interest Rate

[See **Annual percentage rate**]

232. Contract Month

The month in which a futures contract is scheduled to mature by making or accepting a delivery.

233. Contract Specification

The precise definition of the good or service to be delivered in the futures contract.

234. Contractual Institutions

Financial institutions that attract savings from the public by offering contracts that protect the saver against risk in the future, such as insurance policies and pension plans.

235. Contrarian Analysis

Contrarian investing is an investment style in which investors purposefully go against prevailing market trends by selling when others are buying, and buying when most investors are selling. Contrarian investors believe that people who say the market is going up do so only when they are fully invested and have no further purchasing power. At this point, the market is at a peak. So, when people predict a downturn, they have already sold out, and the market can only go up at this point.

236. Contribution Margin

Amount that each additional unit produced, such as a jet engine, contributes to after-tax profit of the whole project: $(\text{Sales price} - \text{Variable cost}) \times (1 - T_c)$, where T_c is the corporate tax rate.

237. Control Theory

Control theory provides a large set of theoretical and computational tools with applications in a wide range of fields, ranging from “pure” areas of mathematics up to applied sciences like finance. Stochastic optimal control is a well-established and important tool of mathematical finance.

238. Control Variate Method

A technique used in Monte Carlo valuation in which simulated asset prices are used to compute two derivatives prices: the price of the derivative that is being valued, and the price of a related derivative for which the value is known. The error in valuing the derivative with a known price is used as a control for that with the unknown price.

239. Controller

The firm’s controller traditionally manages accounting, cost analysis, and tax planning and usually reports to the chief financial officer (CFO).

240. Convenience Yield

A measure of the benefits from ownership of an asset that are not obtained by the holder of a long futures contract on the asset.

241. Conventional Mortgage

A mortgage or deed or trust that is not obtained under a government insured program.

242. Convergence Property

The convergence of futures prices and spot prices at the maturity of the futures contract.

243. Conversion

A risk-free position consisting of an asset, a purchased put, and a written call. For example, we can create a risk-free position by buying a stock and a put of the stock and sell the call of the stock to create a Treasury bill.

244. Conversion Factor

A factor used to determine the number of bonds that must be delivered in the Chicago Board of Trade bond futures contract. For example, the Treasury bond futures contract allows a party with the short position to choose to deliver any bond that has a maturity of more than 15 years and that is not callable within 15 years. When a particular bond is delivered, a parameter known as its *conversion factor* defines the price received by the party with the short position.

245. Conversion Fee

Fee charged for converting a loan commitment to a term loan.

246. Conversion Premium

Difference between the conversion price and the current stock price divided by the current stock price. [See **Conversion price**]

247. Conversion Price

The conversion price, in general, equals:

$$\text{Conversion price} = \frac{\text{Par value of bond}}{\text{Conversion ratio}}$$

It implies that the amount of par value exchangeable for one share of common stock.

248. Conversion Ratio

The number of shares per \$1,000 bond (or debenture) that a bondholder would receive if the bond were converted into shares of stock. [See **Convertible bonds**]

249. Conversion Value

What a convertible bond would be worth if it were immediately converted into the common stock at the current price. [See **Convertible bonds**]

250. Convertibility

A feature of some preferred stocks and bonds which entitles the holder to exchange those securities for a specific number of shares of common stock.

251. Convertible Bonds

Convertible bonds may vary the amount of money the bondholder ultimately receives. A convertible bond can be converted, as the investor's option, into a specified number of shares of the issuer's common stock (defined as the bond's **conversion ratio**). The conversion ratio is set to make conversion unattractive initially. If the firm meets with success, however, its stock price will rise, and the bond's price will be affected by its **conversion value** (the stock price times the conversion ratio) rather than just its value as a straight bond.

252. Convertible Debt

A bond that may be exchanged for common stock in the same firm. [See **Convertible debt**]

253. Convertible Risk

The variability of return causes when one type of security is converted into another type of security. If a bond or a preferred stock is convertible into a stated number of shares of common stock of the corporation issuing the original security, the rate of return of the investment may vary because the value of the underlying common stock has increased or decreased. A convertible security normally has a lower coupon rate, or stated dividend (in the case of preferred stocks), because investors are willing to accept a lower contractual return from the company in order to be able to share in any rise in the price of the firm's common stock.

254. Convertible Securities

A convertible security is a bond or preferred stock issue that typically gives its holder the right to exchange it for a stipulated number of shares of common stock of the issuing corporation during a specified period of time. Therefore, convertible bonds and convertible preferred stock represent options to the security holder. If the price of common stock rises sufficiently, holders of these securities will find it profitable to exercise their conversion rights. As for a warrant, such a right will have some positive value in the market, so the market will accept a lower coupon rate

on the corporation's convertible bonds than it would demand for a bond with no conversion privilege.

Convertible bonds are especially attractive when management prefers to raise capital by issuing equity rather than debt, but believes that transient influences have led the market to temporarily undervalue its common stock. [See also **Convertible bonds**] If this perception is correct, the stock price will rise and, as a result, debt will be converted to equity. A convertible bond issue may offer an advantage over a bond issue with warrants since managers can predict how much capital the issue will raise.

The exercise of a warrant raises further capital for the firm; conversion simply substitutes equity for debt. The conversion of a bond issue for common stock does not raise new capital, but it does implicitly increase cash flow if the conversion occurs prior to the bond's maturity date, by reducing future coupon payments.

A further distinction between warrants and convertible bonds is that warrants are not callable, while the issuer generally can call a convertible bond. The bondholder can be offered the option of converting it within a short time period or surrendering it at a specific cash price. As with all callable bonds, investors demand higher returns for callable, convertible securities. Firms are willing to pay this higher price in exchange for management flexibility.

We have seen why a corporation might want to issue a hybrid security rather than straight debt and/or equity. What about the investor? These securities may be particularly attractive when investors have trouble assessing the riskiness of a corporation's future business activities. If the corporation embarks on a high-risk enterprise, holders of straight bonds will be in the unappealing position of gaining nothing if the enterprise succeeds and facing greatly increased default risk if it fails.

Warrants or conversion privileges can restore some balance. By exercising a warrant or converting a bond to stock, the bondholder can share in any success resulting from a risky venture. This reduces the importance of assessing the future business risk of a corporation's activities.

255. Convex

Convex shaped like the cross section of a bowl. Convex with shapes such that of price-yield relationship are said to be convex and curvature of the price yield curve is called the convexity of the bond.

256. Convexity and Duration

Duration appears a better measure of a bond's life than maturity because it provides a more meaningful relationship with interest-rate changes. This relationship has been expressed by Hopewell and Kaufman (1973) as:

$$\frac{\Delta P}{P} = \frac{D}{(1+r)} \Delta r = D \times \Delta r, \quad (3)$$

where Δ = “change in”; P = bond price; D = duration; and r = market interest rate or bond yield.

For example, a bond with 5 years’ duration will decline in price by approximately 10% when market yield increases by 2%. Note that $\Delta P/P$ on the left-hand side of Eq. 3 is the percentage change in price, and Δr is the absolute change in yield level and not percentage change.

The duration rule in Eq. 3 is a good approximation for small changes in bond yield, but it is less accurate for large changes. Equation 3 implies that percentage change in bond price is linearly related to change in yield to maturity. If this linear relationship is not hold, then Eq. 4 can be generalized as

$$\frac{\Delta P}{P} = -D \times \Delta r + 0.5 \times \text{Convexity} \times (\Delta r)^2 \quad (4)$$

The *Convexity* is the second derivative of a bond’s price with respect to a change in the interest rate, divided by the bond price. In other words, it refers to the degree of curvature of the price-yield curve around some interest level. [See **Convex**] The *Convexity* in Eq. 4 is the rate of change of the slope of the price-yield curve as follows

$$\begin{aligned} \text{Convexity} &= \frac{1}{P} \times \frac{\partial^2 P}{\partial r^2} = \frac{1}{P \times (1+r)^2} \sum_{t=1}^T \left[\frac{C_t}{(1+r)^t} (t^2 + t) \right] \\ &\approx 10^8 \left[\frac{\Delta P_-}{P} + \frac{\Delta P_+}{P} \right] \end{aligned} \quad (5)$$

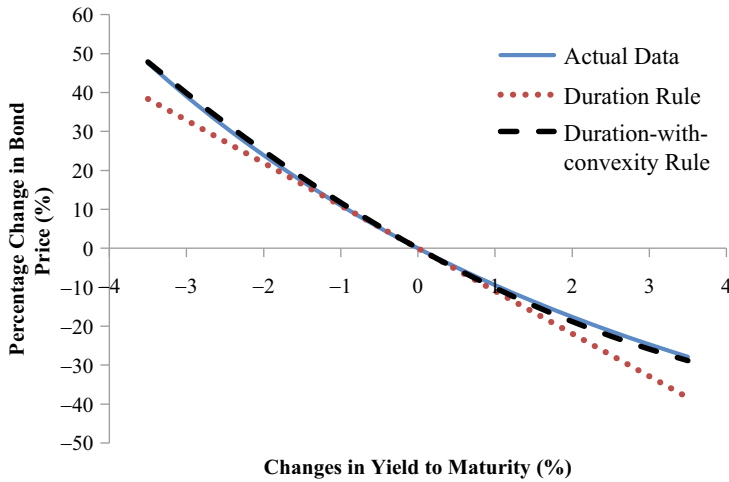
where C_t is the cash payment in period t ; T is the maturity; C_t represents either a coupon payment before maturity or final coupon plus par value at the maturity date. ΔP_- is the capital loss from a one-basis-point (0.0001) increase in interest rates and ΔP_+ is the capital gain from a one-basis-point (0.0001) decrease in interest rates. This approximation of convexity is referred by Saunders and Cornett (2011).

In Eq. 4, the first term of on the right-hand side is the same as the duration rule, Eq. 3. The second term is the modification for convexity. Notice that for a bond with positive convexity, the second term is positive, regardless of whether the yield rises or falls.

The more accurate Eq. 4, which accounts for convexity, always predicts a higher bond price than Eq. 3. Of course, if the change in yield is small, the convexity term, which is multiplied by $(\Delta r)^2$ in Eq. 4, will be extremely small and will add little to the approximation. In this case, the linear approximation given by the duration rule will be sufficiently accurate. Thus convexity is more important as a practical matter when potential interest rate changes are large.

The following figure is drawn by the assumptions that the bond with 20-year maturity and 7.5% coupon sells at an initial yield to maturity of 7.5%. Because the coupon rate equals yield to maturity, the bond sells at par value, or \$1,000. The modified duration and convexity of the bond are 10.95908 and 155.059 calculated by the definition [See **Duration**] and the approximation formula in Eq. 5, respectively.

The Relationship between Percentage Changes in Bond Price and Changes in YTM



If the bond's yield increases from 7.5% to 8.0% ($\Delta r = 0.005$), the price of the bond actually falls to \$950.9093. Based on the duration rule, the bond price falls from \$1,000 to \$945.2046 with a decline of 5.47954% Eq. 3 as follows:

$$\begin{aligned}\frac{\Delta P}{P} &= -D \times \Delta r = -10.95908 \times 0.005 \\ &= -0.0547954, \text{ or } -5.47954\%\end{aligned}$$

If we use Eq. 4 instead of Eq. 3, we get the bond price falls from \$1,000 to \$947.1482 with a decline of 5.28572% by Eq. 4:

$$\begin{aligned}\frac{\Delta P}{P} &= -D \times \Delta r + 0.5 \times \text{Convexity} \times (\Delta r)^2 \\ &= -10.95908 \times 0.005 + 0.5 \times 155.059 \times (0.005)^2 \\ &= -0.0528572, \text{ or } -5.28572\%\end{aligned}$$

The duration rule used by Eq. 3 is close to the case with accounting for convexity in terms of Eq. 4.

However, if the change in yield are larger, 3% ($\Delta r = 0.03$), the price of the bond actually fall to \$753.0727 and convexity becomes an important matter of pricing the percentage change in bond price. Without accounting for convexity, the price of the bond on dash line actually falls from \$1,000 to \$671.2277 with a decline of 32.8772% based on the duration rule, Eq. 3 as follows:

$$\begin{aligned}\frac{\Delta P}{P} &= -D \times \Delta r = -10.95908 \times 0.03 \\ &= -0.328772, \text{ or } -32.8772\%\end{aligned}$$

According to the duration-with-convexity rule, Eq. 4, the percentage change in bond price is calculated in the following equation:

$$\begin{aligned}\frac{\Delta P}{P} &= -D \times \Delta r + 0.5 \times \text{Convexity} \times (\Delta r)^2 \\ &= -10.95908 \times 0.03 + 0.5 \times 155.059 \times (0.03)^2 \\ &= -0.258996, \text{ or } -25.8996\%\end{aligned}$$

The bond price \$741.0042 estimated by the duration-with-convexity rule is close to the actual bond price \$753.0727 rather than the price \$671.2277 estimated by the duration rule. As the change in interest rate becomes larger, the percentage change in bond price calculated by Eq. 3 is significantly different from it calculated by Eq. 4. Saunders and Cornett (2011) have discussed why convexity is important in the risk management of financial institutions.

257. Convolutional Neural Networks (CNNs)

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on the shared-weight architecture of the convolution kernels that scan the hidden layers and translation invariance characteristics.

258. Copula Function

A copula function is simply a specification of how the univariate marginal distributions combine to form a multivariate distribution. For example, if we have N -correlated uniform random variables, $U_1, U_2, \dots, U_N$, then

$$C(u_1, u_2, \dots, u_N) = P_r\{U_1 < u_1, U_2 < u_2, \dots, U_N < u_N\}$$

is the joint distribution function which gives the probability that all of the uniforms are in the specified range.

In a similar manner we can define the Copula function for the default times of N assets

$$\begin{aligned}C(F_1(T_1), F_2(T_2), \dots, F_N(T_N)) &= P_r\{U_1 < F_1(T_1), U_2 < F_2(T_2), \dots, \\ &\quad U_N < F_N(T_N)\}\end{aligned}$$

where $F_i(T_i) = P_r\{t_i < t\}$

David X. Li (2000) has shown how copula function can be used to estimate default correlation. [See **Default correlation**]

259. Core Capital

Tier 1 capital consisting primarily of stockholder's equity.

260. Core Deposits

A base level of deposits a bank expects to remain on deposit, regardless of the economic environment.

261. Correspondent Bank

A bank that provides services, typically check clearing, to other banks.

262. Corporate Bonds

Long-term debt issued by private corporations typically paying semiannual coupons and returning the face value of the bond at maturity.

263. Corporate Governance

Corporate governance is the system of rules, practices, and processes by which a firm is directed and controlled. Corporate governance essentially involves balancing the interests of a company's many stakeholders, such as shareholders, senior management executives, customers, suppliers, financiers, the government, and the community. Since corporate governance also provides the framework for attaining a company's objectives, it encompasses practically every sphere of management, from action plans and internal controls to performance measurement and corporate disclosure.

264. Corporate Leverage

Corporate leverage is used to refer to the debt floated by the corporation.

265. Corporate Note

A debt contract (IOU) of a corporation whose original maturity is 5 years or less.

266. Corporations

Proprietorships are the most numerous form of business organization, but in terms of market value, corporations are the dominant form. A corporation is a legal person in the eyes of the law, separate in concept from its owners and managers. As a person, it has rights, duties, privileges, and obligations.

The corporate organizational form has several advantages. As a separate legal entity, its life does not depend on that of its owners. Unlike a proprietorship or partnership, the death of a shareholder does not force the corporation to stop doing business. Shares of ownership in the corporation, especially those listed on stock exchanges such as the New York Stock Exchange, can be traded at easily discernible prices. Issuing shares gives a corporation access to much larger pools of capital than a partnership or proprietorship. As a legal entity, it can borrow money in its own name. Also, as owners of a corporation have limited liability, the most they can lose is their investment.

A major disadvantage of the corporate organizational form is the taxation of earnings. Depending upon the income level, corporate income may be taxed at higher rates than proprietor or partnership income. In addition, corporate dividends are taxed twice. As corporations pay dividends from after-tax earnings, they represent funds that have been taxed once at the corporate level. Investors then pay taxes on these dividends again, as part of their personal income.

Two special forms of corporate organization in the United States allow dividends to escape double taxation. [See **Subchapter S corporation** and **Limited liability company**]

Many countries' laws recognize the corporate form of organization. US corporations may use the suffixes "Inc." or "Corp." to designate themselves. British corporations use the suffix "PLC," for *public limited company*, in which *limited* refers to shareholders' liability in the firm. The suffix "AG" following the same names of firms in Germany, Austria, Switzerland, or Liechtenstein is an abbreviation for *Aktiengesellschaft*, which means *corporation*. Some countries allow corporations to sell bearer shares, which allow the owners to remain anonymous. A history of social upheavals, wars, and high taxation in Europe led to the evolution of bearer shares to allow owners to remain anonymous and thus escape taxation from their governments or identification if their governments were overthrown. Suffixes of "NV" (*Naamioze Venootschap*) in the Netherlands and "SA" (*Societe Anonyme*) in France and Belgium designate such firms.

In sum, a corporation is one type of business organization that is created as a distinct "legal person" composed of one or more actual individuals or legal entities, primary advantages of a corporation include limited liability, ease of ownership, transfer, and perpetual succession.

267. Correlation

Correlation is a statistical concept that relates movements in one set of variables to movements in another. **Covariance** can indicate a positive, zero, or negative relationship between two variables, but little else. [See also **Covariance**] Correlation, however, shows the *strength* of the linear relationship between two sets of variables. The correlation coefficient between two set of numbers, denoted by the small Greek letter rho (ρ), is computed as:

$$\rho_{12} = \frac{\text{cov}(R_1, R_2)}{\sigma_1 \sigma_2}$$

where σ_1 and σ_2 are the standard deviations of the two number series. Mathematically, the correlation will always lie between -1.0 and $+1.0$, inclusively. As correlation approaches $+1.0$, it indicates a stronger *positive* linear relationship between the two series of numbers. As the correlation approaches -1.0 , it indicates a stronger *negative* linear relationship between the two series. The greatest reduction in risk occurs when two strongly negatively correlated assets are placed in the same portfolio. Correlations close to zero represent weak linear relationship; a correlation of zero implies that no linear relationship exists.

268. Correlation Coefficient

A statistic in which the covariance is scaled to a value between -1 (perfect negative correlation) and $+1$ (perfect positive correlation). [See **Correlation**]

269. Cost of Carry

The interest cost of owning an asset, less lease or dividend payments received as a result of ownership; the net cash flow resulting from borrowing to buy an asset.

270. Cost of Common Equity

Unlike debt and preferred stock, cash flows from common equity are not fixed or known beforehand, and their risk is harder to evaluate. In addition, firms have *two* sources of common equity – retained earnings and new stock issues – and thus *two* costs of common equity. It may be clear that there is an explicit cost (i.e., dividends and flotation costs) associated with issuing new common equity. But while the firm pays no extra dividends or flotation costs to use retained earnings, their use is not free; we must consider the opportunity cost of using money that could have been distributed to shareholders.

Retained earnings represent the portion of net income that the firm does not distribute as dividends. From the shareholders' perspective, the opportunity cost of retained earnings is the return the shareholders could earn by investing the funds in assets whose risk is similar to that of the firm. To maximize shareholder wealth, management must recognize that retained earnings have a cost. That cost, k^{re} , is the return that shareholders expect from their investment in the firm.

271. Cost of Debt

The firm's unadjusted cost of debt financing equals the yield to maturity (YTM) on new debt, either a long-term bank loan or a bond issue. The yield to maturity represents the effective annual rate or cost to the firm of borrowing funds in the current market environment. Coupon rates from previously issued bonds reveal little about the firm's present financing costs. The firm's *current* financing costs determine its current cost of capital.

A firm can determine its cost of debt by several methods. If the firm targets an “A” rating (or any other bond rating), a review of the yields to maturity on A-rated bonds can provide an estimate of the firm’s current unadjusted borrowing costs. Several additional factors will affect the firm’s specific borrowing costs, including covenants and features of the proposed bond issue as well as the number of years until the bond or loan matures or comes due. It is important to examine bonds whose ratings and characteristics resemble those the firm wants to match.

In addition, the firm can solicit the advice of investment bankers on the cost of issuing new debt. Or, if the firm has debt currently trading, it can use public market prices and yields to estimate its current cost of debt. The publicly traded bond’s yield to maturity can be found using the techniques for determining the return on an investment. Finally, a firm can seek long-term debt financing from a bank or a consortium of banks. Preliminary discussions with the bankers will indicate a ballpark interest rate the firm can expect to pay on the amount it borrows. (For calculation of YTM see **Yield to maturity**)

272. Cost of Equity Capital

The required return on the company’s common stock in capital markets. It is also called the equity holders’ required rate of return because it is what equity holders can expect to obtain in the capital market. It is a cost from the firm’s perspective. [See **Cost of common equity**]

273. Cotango

An increment added to a futures price to cover the carrying costs until delivery occurs at the schedule settlement date. (Also called Forwardation.) Therefore, the futures price must exceed the expected future spot price.

274. Counterparties

The buyer and seller of a derivative such as swap are counterparties. Usually, not always, a financial institution serves as an intermediary between the counterparties. When bank and the company agree on an at-the-money forward exchange contract or swap, the company is at risk if the bank fails, just as much as the bank is at risk if the counterparty fails. After inception, swap positions often move in/out-of-the-money and the relative credit risk changes accordingly. [See also **Interest rate swap**]

275. Country Risk

The credit risk that government or private borrowers in a specific country will refuse to repay their debts as obligated for other than pure economic reasons. For example, the repayments from foreign borrowers may be interrupted because of interference from foreign government.

276. Country Selection

A type of active international management that measures the contribution to performance attributable to investing in the better-performing stock markets of the world.

277. Coupon

The stated interest on a debt instrument. In bonds, notes, or other fixed income securities, the stated percentage rate of interest, usually paid twice a year.

278. Coupon Bond

A security that obligates the issuer to make interest payments called coupon payments over the life of the bond, then to repay the face value at maturity.

279. Coupon Effect

The size of a debt security's promised interest rate (coupon) influences how rapidly its price moves with changes in market interest rates.

280. Coupon Interest Rate (Coupon Rate)

The coupon interest rate is the percentage of the par value to be paid annually, as interest, to the bond holder.

281. Coupon-Reinvestment Risk

It is the expected yield calculated by assuming that all coupon cash flows would be reinvested at the same yield that exists at the time of purchase. If rates began to fall, it would be impossible to reinvest the coupon at a rate high enough to produce the anticipated yield. If rates increase, the coupon cash flow will be reinvested at a higher rates and produce a return above expectation.

282. Covariance

Covariance is a statistical concept that relates movements in one set of variables to movements in another. For example, the covariance between two sets of returns, R_1 and R_2 , is:

$$\text{cov}(R_1, R_2) = \frac{\sum_{t=1}^N (R_{1t} - \bar{R}_1)(R_{2t} - \bar{R}_2)}{N - 1}$$

where N represents the number of joint observations of the assets' returns; R_{1t} and R_{2t} represent the t th observations of R_1 and R_2 ; and $\bar{R}_1$ and $\bar{R}_2$ represent the average returns on the two assets.

A negative covariance produces a situation where when one set of returns is rising, the other is usually falling, and vice versa. A zero covariance means that the two time series have no linear relationship.

Covariance can indicate a positive, zero, or negative relationship between two variables but little else. To look at the *strength* of the linear relationship between two sets of variables we look at the **correlation**. [See also **Correlation**]

283. Covenants

A bond indenture may include covenants, which can impose restrictions or extra duties on the firm. Covenants are most effective when they are specific measures that state the acceptable limits for change in the obligor's financial and overall condition. They clearly define what is meant by "significant" deterioration in the obligor's credit quality. Financial covenants are more explicit (and therefore more desirable) than a "material adverse change" clause. Cross default provisions are common: allowing acceleration of debt repayment. These provisions affect the credit rating of the issue and the firm's financing costs. Restrictive covenants designed to protect bondholders and maintain the value of their investment can reduce the issuer's financing costs. The firm must decide if the restrictions and duties are worth the access to lower cost funds. The trustee ensures that the issuer observes any bond covenants. If the issuer violates a covenant, the issue is technically in default and the trustee can pursue a legal remedy, including immediate redemption of the bondholders' principal, in court.

Covenants are an example of a mechanism to control bondholder-firm agency problems. The covenants help ensure that management's actions do not unduly jeopardize the firm's liquidity and the bondholders' security. Examples of covenants include stipulations that the firm must maintain a minimum level of net working capital, maintain a minimum interest coverage ratio, keep pledged assets in good working order, and send audited financial statements to bondholders. Other examples include restrictions on the amount of the firm's debt, its dividend payments, and asset sales.

284. Coverage Ratios

Ratios of company earnings to fixed costs. Low or falling coverage ratios signal possible cash flow difficulties. [See **Interest coverage ratio**]

285. Covered Call

Covered call is a long position in an asset together with a written call on the same asset. Covered calls are far less risky than naked calls, because the worst that can happen is that the investor is required to sell shares already owned at a below their market value.

286. Covered Interest Arbitrage

A zero-investment strategy with simultaneous borrowing in one currency, lending in another, and entering into a forward contract to guarantee the exchange rate when the loans mature.

287. Covered Write

A long position in an asset coupled with sale of a call option on the same asset.

288. Cox-Ingersoll-Ross (CIR) Model

The Cox-Ingersoll-Ross (CIR) model is a mathematical formula used to model interest rate movements and is driven by a sole source of market risk. It is used as a method to forecast interest rates and is based on a stochastic differential equation.

289. Crack Spread

Crude oil is generally refined to make petroleum products, in particular heating oil and gasoline. The split of oil into heating oil and gasoline can be complemented by a process known as “cracking.” Hence, the difference between the price of crude oil futures and that of equivalent amounts of heating oil and gasoline is known as crack spread.

290. Creative Financing

Innovative financing techniques used to make single-family homes more affordable for the average home buyer.

291. Credit

A loan of funds in return for a promise of future payment.

292. Credit Analysis

Credit analysis is a type of financial analysis that an investor or bond portfolio manager performs on companies, governments, municipalities, or any other debt-issuing entities to measure the issuer’s ability to meet its debt obligations. Credit analysis seeks to identify the appropriate level of default risk associated with investing in that particular entity’s debt instruments.

293. Credit Bureau

An association that collects and provides information on the credit (payment) histories of borrowers.

294. Credit Card

A plastic card (sometimes equipped with a microprocessor) that allows the holder to borrow cash or to pay for goods and services with credit.

295. Credit Card Delinquency

A consumer is characterized as being delinquent when they have fallen behind on their payment obligations for a monthly bill. When it comes to credit cards, delinquency does not mean that you have failed to pay off your full balance in a given month, but rather that you did not make the required minimum payment.

296. Credit Check

Efforts by a lender to verify the accuracy of information provided by potential borrowers.

297. Credit Department

The bank department where credit information is collected and analyzed to make credit decisions.

298. Credit Default Swaps

A credit default swap (CDS) is a financial derivative or contract that allows an investor to “swap” or offset his or her credit risk with that of another investor. For example, if a lender is worried that a borrower is going to default on a loan, the lender could use a CDS to offset or swap that risk. To swap the risk of default, the lender buys a CDS from another investor who agrees to reimburse the lender in the case the borrower defaults. Most CDS will require an ongoing premium payment to maintain the contract, which is like an insurance policy. A credit default swap is the most common form of credit derivative and may involve municipal bonds, emerging market bonds, mortgage-backed securities, or corporate bonds.

299. Credit Derivatives

A claim where the payoff depends upon the credit rating or default status of a firm. These include credit options, credit swaps, credit forwards, and others.

300. Credit Enhancement

A guarantee or letter of credit backing for a loan which improves the creditworthiness of the contract.

301. Credit Exposure

The amount subject to changes in value upon a change in credit quality through either a market-based revaluation on the event of an up (down) grade or the application of a recovery fraction in the event of default.

302. Credit File

Information related to a borrower's loan request, including application, record of past performance, loan documentation, and analyst opinions.

303. Credit Instrument

Device by which a firm offers credit, such as an invoice, a promissory note, or a conditional sales contract.

304. Credit Limit

The maximum amount that a borrower is allowed to borrow against a loan commitment or credit line.

305. Credit-Linked Notes (CLNs)

Bonds that have payments determined at least in part by credit events (e.g., default) at a different firm. These also refer to asset-backed securities which were issued against the loan portfolio.

Credit-linked notes exist in a number of forms, but all of them contain a link between the return they pay and the credit-related performance of the underlying asset. A standard CLN is a security, usually issued by an investment-graded entity that has an interest payment and fixed maturity structure similar to a vanilla bond. The performance of the CLN, however, including the maturity value, is linked to the performance of a specified underlying asset or assets as well as that of the issuing entity. CLNs are usually issued at par. They are often used as a financing vehicle by borrowers in order to hedge against credit risk; CLNs are purchased by investors to enhance the yield received on their holdings. Hence, the issuer of the CLN is the protection buyer and the buyer of the note is the protection seller.

306. Credit Period

Time allowed a credit purchaser to remit the full payment for credit purchases. Credit periods vary among different industries. For example, a jewelry store may sell diamond engagement rings for 5/30, net 4 months (the company require final payment within 4 months but offer a 5% discount to customers who pay within 30 days). A food wholesaler, selling fresh fruit and produce, might use net 7. Generally, a firm must consider three factors in setting a credit period. (1) The probability that the customer will not pay. A firm whose customers are in high-risk businesses may find itself offering restrictive credit terms. (2) The size of the account. If the account is small, the credit period will be shorter. Small accounts are more costly to manage, and small customers are less important. (3) The extent to which the goods are perishable. If the collateral values of the goods are low and cannot be sustained for long periods, less credit will be granted.

307. Credit Quality

Generally meant to refer to an obligor's relative chance of default, usually expressed in alphabetic terms (e.g., Aaa, Aa, A). In credit metrics analysis, the credit quality includes also the volatility of up (down) grades.

308. Credit Rating

A measure of the creditworthiness of a bond issue. In addition to bond credit rating, there is a growing trend toward the "credit rating" of loans offered for sale. Unlike bonds, a loan credit rating reflects more than the financial soundness of the underlying borrowing corporation. In particular, the value of the underlying collateral can change a loan's credit rating up to one full category above a standard bond rating. As more loans are rated, their attractiveness to secondary market buyers is likely to increase.

309. Credit Ratings Transition Matrix

A table showing the probability that a company will move from one credit rating to another during a certain period of time.

310. Credit Risk

The cash flows to be received by bond market investors are not certain; like individuals, corporate debtors may pay interest payments late or not at all. They may fail to repay principal at maturity. To compensate investors for this credit or **default risk**, rates of return on corporate bonds are higher than those on government securities with the same terms of maturity. Government securities are presumed to be free of credit risk. Generally, as investors perceive a higher likelihood of default, they demand higher default-risk premiums. Since perceptions of a bond's default risk may change over its term, the bond's yield to maturity also may change, even if all else remains constant.

311. Credit Risk Classification

Credit risk can be analyzed and measured from different perspectives. A default-mode valuation considers all the three risks, namely, default risk, recovery risk, and exposure risk. A key concept of credit risk measurement is "expected loss": it is the average loss generated in the long run by a group of credit facilities. The approach to determine expected loss may be financial or actuarial. The availability of information such as expected and unexpected losses can substantially innovate the way credit strategies are set. By providing evidence on the relative risk adjusted profitability of customer relationships, economic capital can be used in optimizing the risk-return trade-off in bank portfolios.

312. Credit Risk in Banking

Credit risk is most simply defined as the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms. The goal of credit risk management is to maximize a bank's risk-adjusted rate of return by maintaining credit risk exposure within acceptable parameters. Banks need to manage the credit risk inherent in the entire portfolio as well as the risk in individual credits or transactions. Banks should also consider the relationships between credit risk and other risks. The effective management of credit risk is a critical component of a comprehensive approach to risk management and essential to the long-term success of any banking organization.

313. Credit Scoring

The use of a statistical model based on applicant attributes to assess whether a loan automatically meets minimum credit standards. The model assigns values to potential borrowers' attributes, with the sum of the values compared to a threshold. More specifically, this is a reference to the application of linear discriminant analysis to combine financial ratios to quantitatively predict the relative chance of default. [See **Credit scoring model**]

314. Credit Scoring Model

Using financial ratio analysis to evaluate credit risk is certainly helpful. Yet, the decision that must be made following the examination of such data can be complicated by the difficulty of interpreting conflicting ratios. Different ratios often imply different predictions for the same firm. To overcome such ambiguity, information from several financial ratios can be combined into a single index. The resulting multivariate financial model will yield a single number for classifying the firm in terms of credit risk. For example, the multivariate financial model developed by Professor Altman's (1968, pp. 589–609), which is as follows:

$$Y_i = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5,$$

where:

X_1 = Working capital/Total assets

X_2 = Retained earnings/Total assets

X_3 = EBIT/Total assets

X_4 = Market value of equity/Book value of total debt

X_5 = Sales/Total assets

By substituting the financial ratio information for an individual company into this model, we can obtain financial Z-scores. This financial Z-score can be used to determine financial condition of a firm.

315. Credit Sensitive Notes

The coupon rates on credit sensitive notes increase (or decrease) if the issuer's bond rating falls (or rises). This compensates investors for changes in the issuer's credit quality over the life of the note.

316. Credit Spread

A credit spread is the difference in yield between a US Treasury bond and another debt security of the same maturity but different credit quality. Credit spreads between US Treasuries and other bond issuances are measured in basis points, with a 1% difference in yield equal to a spread of 100 basis points. As an example, a 10-year Treasury note with a yield of 5% and a 10-year corporate bond with a yield of 7% are said to have a credit spread of 200 basis points. Credit spreads are also referred to as "bond spreads" or "default spreads." Credit spread allows a comparison between a corporate bond and a risk-free alternative. A credit spread can also refer to an options strategy where a high premium option is written, and a low premium option is bought on the same underlying security. This provides a credit to the account of the person making the two trades.

317. Credit-Scoring Model

Credit scoring is a statistical analysis performed by lenders and financial institutions to determine a person's or a small, owner-operated business' creditworthiness. Credit scoring is used by lenders to help decide whether to extend or deny credit. A credit score can impact many financial transactions, including mortgages, auto loans, credit cards, and private loans. Credit scoring models may differ slightly in how they score credit. Fair Isaac Corporation's credit scoring system, known as a FICO score, is the most widely used credit scoring system in the financial industry, employed by more than 90% of top lenders.

318. Credit Spread Option

Option whose payoff depends on the spread between the yields earned on two assets. Options can be written on many spread: bond spreads, credit default swap spreads, and asset swap spreads.

319. Credit Union

A nonprofit organization that offers financial services to qualifying members. Credit unions do not pay state and federal income taxes and thus operate at a competitive advantage to other depository institutions.

320. Credit Value at Risk

The credit loss that will not be exceeded at some specified confidence level. [See **Value at risk**]

321. Credit Metrics Model

Credit metrics was introduced in 1997 by J. P. Morgan and its co-sponsors (Bank of America, Union Bank of Switzerland, etc.) as a value at risk (VAR) framework to apply to the valuation and risk of nontradable assets such as loans and privately placed bonds. Thus, while Risk Metrics seeks to answer the question “If tomorrow is a bad day, how much will I loose on tradable assets such as stocks and bonds?” Credit Metrics asks, “If next year is a bad year, how will I loose on my loans and loan portfolio.”

322. Creditor

Person or institution that holds the debt issued by a firm or individual. Bond holder is the creditor; creditors do not usually have voting power. The device used by creditors to protect themselves is the loan contracts (i.e., the indenture).

323. Creditors' Committee

A method of adjusting a capital structure without bankruptcy proceedings involves the operation of the enterprise by a group of creditors, called a creditors' committee. These representatives manage the firm until it gathers sufficient liquid capital to satisfy existing claims or until an acceptable composition is found.

There is no legal compulsion for any creditor to accept an out-of-court adjustment. Any creditor can delay the process if it is dissatisfied with a proposal by the majority (or minority) of creditors to relieve the financial burden on the firm. The unhappy creditor can refuse the arrangement and insist that a claim be met in full; if it is not, the creditor can take the firm to court to be liquidated or reorganized.

324. Cross Rate

An exchange rate may be quoted as a cross-rate, the rate of a non-US dollar currency expressed in terms of another non-US dollar currency.

325. Cross Hedge

Use of a futures contract for a specific asset that differs from the cash asset being hedged. For example, to use index futures to hedge 10-year US government bond. For example IBM holds stocks of GM.

326. Cross Holdings

One corporation holds shares in another firm.

327. Cross-Sectional Analysis

Financial ratios can be used in *cross-sectional analysis*, in which different firms are compared at the same point in time. The best information source for cross-

sectional analysis of firm ratios is the firm's financial statements and their footnotes. These materials appear in annual reports as well as 10-Q and 10-K filing with the Securities and Exchange Commission.

328. Crown Jewels

An anti-takeover tactic in which major assets – the crown jewel – are sold by a firm when faced with a takeover threat. This is sometimes referred to as the scorched earth strategy

329. Crush Spread

Soybean generally can be crashed to produce soybean meal and soybean oil. Therefore, the difference between the price of a quantity of soybeans and that of the soybean meal and oil that can be produced by those soybeans.

330. Cum Dividend

With dividend before the ex-dividend date the stock is said to trade cum dividend.

331. Cumulative Abnormal Return (CAR)

Sum of differences between the expected return on a stock and the actual return that comes from the release of news to the market. The abnormal return on a given stock for a particular day can be calculated by subtracting the market's return on the same day (R_m) – as measured by a broad based index such as the S&P composite index – from the actual return (R) on the stock for the day. $AR = R - R_m$. Cumulative abnormal return (CAR) is the total abnormal return for the period surrounding an announcement on the release of information. The CAR generally can be measured by cumulative average residual. [See **Cumulative average residual**]

332. Cumulative Average Residual (CAR)

Following Fama, Fisher, Jensen, and Roll (1969, pp. 1–21), the cumulative average residual (CAR) can be defined as follows:

$$CAR = \sum_{t=1}^T AR_t$$

where:

$$AR_t = \frac{1}{N} \sum_{j=1}^N e_{jt}$$

$$e_{jt} = R_{jt} - \beta_j R_{mt}$$

R_{jt} = rate of return for security j in period t

β_j = beta coefficient for j th security

R_{mt} = market rate of return in period t

T = the number of months being summed ($T = 1, 2, \dots, M$)

N = the total number of months in the sample

333. Cumulative Distribution Function

A function giving the probability that a value drawn from a distribution will be less than or equal to some specified value.

334. Cumulative Dividend

Dividend on preferred stock that takes priority over dividend payments on common stock. Dividends may not be paid on the common stock until all past dividends on the preferred stock have been paid.

335. Cumulative Normal Distribution Function

The cumulative distribution function for the normal distribution; $N(x)$ in the Black-Scholes equation. [See **Black-Scholes OPM**]

336. Cumulative Probability

The probability that a drawing from the standardized normal distribution will be below a particular value. For example, the probability for a standardized normal distribution that a drawing will be below 0 is clearly 50% because the normal distribution is symmetric. Using statistical terminology, we say that the cumulative probability of 0 is 50%. Statisticians also say that $N(0) = 50\%$.

337. Cumulative Sum (CUSUM)

In statistical quality control, the CUSUM (or cumulative sum control chart) is a sequential analysis technique that is typically used for monitoring change detection. The primary purpose of a CUSUM control chart is to detect small shifts from the process target. These types of control charts are very useful in monitoring a process for statistical control or to see the impact of your process improvement efforts. You can also increase the sensitivity of these charts by using the zone tests and run tests in addition to the points beyond the control limits test for out of control conditions. Cumulative sum (CUSUM) has been frequently used to analyze the earnings announcement effect on stock price.

338. Cumulative Voting

A procedure whereby a shareholder may cast all of his or her votes for one member of the board of directors. The effect of cumulative voting is to permit

minority participation. If cumulative voting is permitted. The total number of votes that each shareholder may cast is determined first. The number is calculated as the number of shares (owned or controlled) multiplied by the number of directors to be elected. Each shareholder can distribute these votes as he or she wishes over one or more candidates.

339. Currency Futures

Contracts that allow businesses or individuals acquiring or selling a foreign currency to protect against future fluctuations in currency prices by shifting the risk of those price changes to someone else willing to bear that risk.

340. Currency Risk

Currency risk, commonly referred to as exchange-rate risk, arises from the change in price of one currency in relation to another. Investors or companies that have assets or business operations across national borders are exposed to currency risk that may create unpredictable profits and losses. Many institutional investors, such as hedge funds and mutual funds, and multinational corporations use forex, futures, options contracts, or other derivatives to hedge the risk.

341. Currency Selection

Asset allocation in which the investor chooses among investments denominated in different currencies.

342. Currency Swap

In a currency swap, two firms agree to exchange an equivalent amount of two different currencies for a specified period of time. A fixed rate is paid in one currency while a floating rate is paid in another. Currency swap is generally used to hedge the currency interest rate risk.

As an example of a typical currency swap, suppose a German company would like to borrow US dollars to finance a foreign investment, but the firm is not known outside Germany. Similarly, a US firm needs DMs for its German subsidiary, but the cost of borrowing in the United States is cheaper than the cost of borrowing in Germany for this firm. Both firms face a similar problem. They can borrow at favorable rates, but not in the desired currency. In this case, a currency swap presents on solution. A bank acting as an intermediary can bring these two firms together and arrange a swap of deutsche marks for dollars. The German firm agrees to pay the US company principal and interest on its dollar borrowings in the United States, while the US firm agrees to pay the costs of the DM borrowings for its German subsidiaries. Each firm thus obtains the best possible rate and eliminates exposure to exchange rate changes by agreeing to exchange currencies.

343. Currency-Translated Index

An investment in an index denominated in a foreign currency, where the buyer bears both currency and asset risk.

344. Current Account

The difference between imports and exports, including merchandise, services, and transfers such as foreign aid.

345. Current Asset

Asset that is in the form of cash or that is expected to be converted into cash in the next 12 months, such as inventory. Current assets are presented in the balance sheet in order of their accounting liquidity – the ease with which they can be converted to cash at a fair price and the time it takes to do so.

346. Current Exposure

For market-driven instruments, the amount it would cost to replace a transaction today should counterparty default. If there is an enforceable netting agreement with the counterparty, then the current exposure would be the net replacement cost; otherwise, it would be the gross amount. [See **Exposure**]

347. Current Liabilities

Obligations that are expected to require cash payment within 1 year or the operating period, whichever is shorter. The three major items found as current liabilities are account payable; accrued wages; and other expenses payable; and note payable. Also, on the balance sheet, Net working capital = cash + other current assets – current liabilities.

348. Current Ratio

Total current assets divided by total current liabilities, used to measure short-term solvency of a firm. [See **Liquidity ratios**]

349. Current Saving

The change in net worth recorded by a sector unit in the economy over the current time period.

350. Current Yield

A bond's annual coupon payment divided by its price. Differs from yield to maturity.

351. Customer Information File

A record of the services used by each customer.

352. Customer Profitability Analysis

A procedure that compares revenues with expenses and the bank's target profit from a customer's total account relationship.

353. Cyclical Liquidity Needs

An estimate of liquid funds needed to cover deposit outflows or loan demand in excess of trend or seasonal factors.

1.4 D

1. D

Definition of the US money supply that includes the total debt of domestic non-financial sectors, consisting of credit market debt of the US government, state and local governments, and private nonfinancial sectors.

2. Data Mining

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing, and computer processing.

3. Data Snooping

Data dredging (also data fishing, data snooping, data butchery, and p-hacking) is the misuse of data analysis to find patterns in data that can be presented as statistically significant, thus dramatically increasing and understating the risk of false positives. This is done by performing many statistical tests on the data and only reporting those that come back with significant results.

4. Date of Payment

Date that dividend checks are mailed.

5. Date of Record

Date on which holders of record in a firm's stock ledger are designated as the recipients of either dividends or stock rights. Dividends will not be paid to those individuals whose notification of purchase is received by the company after this date.

6. Dates Convention

Treating cash flows as being received on exact dates – date 0, date 1, and so forth – as opposed to the end-of-year convention.

7. Day Count

A convention for quoting interest rates.

8. Day Order

A buy order or a sell order expiring at the close of the trading day.

9. Day Trade

A trade that is entered into and closed out in the same day.

10. Daylight Overdrafts

Bank payments from deposits held at a Federal Reserve Bank or correspondent bank in excess of actual collected balances during a day.

11. Days' Receivables

[See **Average collection period**]

12. Days in Receivables

Average collection period. It measures the average amount of time required to collect an account receivable. Suppose in one company, 80% of its customers take the discounts and pay on day 20; the rest pay on day 60. The average collection period is 28 days ($0.8 \times 20 \text{ days} + 0.2 \times 60 \text{ days}$). It is also referred to as days' sales outstanding.

13. Days Sales Outstanding

Average collection period.

14. De Novo Branch

A newly opened branch.

15. Dealer Market

A market where traders specializing in particular commodities buy and sell assets for their own account. Most debt securities are traded in dealer markets. The many bond dealers communicate with one another by telecommunication equipment-wires, computers, and telephones. Investors get in touch with dealer when they want to buy or sell, and can negotiate a deal. Some stocks are traded in the dealer market. The OTC market is an example.

16. Dealer Paper

Short-term commercial notes sold by borrowing corporations and issued through security dealers who contact interested investors to determine if they will buy the notes.

17. Dealer Reserve

An account established by a bank and dealer used to assign the interest that accrues to dealers as they sell loans to a bank.

18. Debenture

A debenture is an unsecured bond, or a bond that pledges no specific assets as security or collateral. In case of default, debenture holders are treated as general creditors of the firm. The riskiest type of bond is a **subordinated debenture**. [See also **Subordinated debenture**]

19. Debit Card

A plastic card that, when used, immediately reduces the balance in a customer's transactions deposit.

20. Debt

Loan agreement that is a liability of the firm. An obligation to repay a specified amount at a particular time.

21. Debt Capacity

Ability to borrow. The amount a firm can borrow up to the point where the firm value no longer increases. A firm's maximum debt capacity is defined as the point

where the advantage derived from an incremental addition of debt to the firm's capital structure is offsetted by the cost incurred.

22. Debt Displacement

The amount of borrowing that leasing displaces. Firms that do a lot of leasing will be forced to cut back on borrowing.

23. Debt Management Policy

The refunding or refinancing of the federal government's debt in a way that contributes to broad national goals and minimizes the burden of the federal debt.

24. Debt Ratio

Total debt divided by total assets. This ratio is used to determine a firm's capital structure.

25. Debt Securities

Financial claims against the assets of a business firm, individual, or unit of government, represented by bonds, notes, deposits, and other contracts evidencing a loan of money.

26. Debt Service

Interest payments plus repayments of principal to creditors, that is, retirement of debt.

27. Debt-to-Assets Ratio

[See **Capital structure ratios**]

28. Debt-to-Equity Ratio

[See **Capital structure ratios**]

29. Debt-to-Equity Ratio

The debt-to-equity ratio is a variation on the total debt ratio. Its total debt divided by total assets. It is frequently used to calculate unleveraged beta.

30. Debt-Like Signal

Debt signaling is a financial theory that correlates a stock's future performance with any current announcements made regarding its debt. Announcements made

about a company taking debt are typically seen as positive news as it can signal the company is creditworthy and is raising capital for the purposes of growth.

31. Debtor-in-Possession Financing

A loan made to a firm which has filed for Chapter 12 bankruptcy protection.

32. Decision Trees

A graphical representation of alternative sequential decisions and the possible outcomes of those decisions. Decision tree can be used to analyze capital budgeting under uncertainty. It can also be used to analyze the option valuation. [See **Binomial model**]

33. Declaration Date

Date on which the board of directors passes a resolution to pay a dividend of a specified amount to all qualified holders of record on a specified date.

34. Dedicated Capital

Total par value (number of shares issued multiplied by the par value of each share). Also called dedicated value.

35. Dedication Strategy

Cash flow matching on a multiperiod basis is referred to as a dedication strategy. In this case, the manager selects either zero-coupon or coupon bonds that provide total cash flows in each period that match a series of obligations. The advantage of dedication is that it is a once-and-for-all approach to eliminating interest rate risk. Once the cash flows are matched, there is no need for rebalancing. The dedicated portfolio provides the cash necessary to pay the firm's liabilities regardless of the eventual path of interest rates.

36. Deed of Trust

Indenture is sometime referred to as deed of trust. [See **Indenture**]

37. Deep-Discount Bond

A bond issued with a very low coupon or no coupon and selling at a price far below par value. When the bond has no coupon, it is also called a pure-discount or original-issue-discount bond.

38. Deep Learning

Deep learning is a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from the raw input. For example, in image processing, lower layers may identify edges, while higher layers may identify the concepts relevant to a human such as digits, letters, or faces. Most recently, deep learning has been extensively used to analyze asset pricing and option pricing. In addition, this method has recently been used to improve technical analysis.

39. De Facto

Existing in actual fact, although not by official recognition.

40. Defalcation

The misappropriation of funds or property by an individual.

41. Default

The failure to make obligated interest and principal payments on a loan.

42. Default Barrier

Default barrier models provide an elegant solution to the credit default prediction challenge since they link the default event to the point at which some continuously evolving quantity hits a known barrier. In structural models of credit risk (see **Structural Default Risk Models**) the process and the barrier are interpreted in terms of capital structure of the firm as the value of the firm and its liabilities. More generally, one can view the process and the barrier as state variables that need not necessarily be observable.

43. Default Prediction

Default prediction in credit analysis uses financial ratio information to predict the probability of default for a company. For example, financial Z-scores can be used to predict the chance of default.

44. Default Premium

A differential in promised yield that compensates the investor for the risk inherent in purchasing a corporate bond that entails some risk of default.

45. Default Probability (DD)

The likelihood that an obligor or counterparty will encounter credit distress within a given time period. “Credit distress” usually leads to either an omitted delayed payment or distressed exchange which would impair the value to senior unsecured debt holders. Note that this leaves open the possibilities that:

- (a) Subordinated debt might default without impairing senior debt value.
- (b) Transfers and clearing might continue even with a senior debt impairment.

This probability can be either marginal default probability (MDP) or cumulative default probability (CDP). The MDP refers the probability that a borrower will default in any given year. The CDP refers the probability that a borrower will default over a specified multiyear period.

46. Default Risk

The chance that interest or principal will not be paid on the due date and in the promised amount under the loan contract. [See also **Credit risk**]

47. Default Correlation

Default correlation is a measurement of the degree to which default of one asset makes more or less likely the default of another asset. One can think of default correlation as being jointly due to (1) a macroeconomic effect which tends to tie all industries into the common economic cycle, (2) a sector-specific effect, and (3) a company-specific effect.

48. Default Probability Density

Measures the unconditional probability of default in a future short period of time. As the asset value of a firm increases, the firm is more likely to remain solvent, the default probability drops.

49. Default Swap

A contract in which the swap buyer pays a regular premium; in exchange, if a default in a specified bond occurs, the swap seller pays the buyer the loss due to the default.

50. Defeasance

A debt-restructuring tool that enables a firm to remove debt from its balance sheet by establishing an irrevocable trust that will generate future cash flows sufficient to service the decreased debt.

51. Deferred Annuities

Tax-advantaged life insurance product. Deferred annuities offer deferral of taxes with the option of withdrawing one's funds in the form of a life annuity.

52. Deferred Availability Credit Items

Checks received for collection for which a bank has not provided credit to the depositor.

53. Deferred Call

A provision that prohibits the company from calling the bond before a certain date. During this period the bond is said to be call protected.

54. Deferred Down Rebate Option

A deferred rebate option for which the current stock price is above the rebate barrier. The payoff to this claim does not depend upon a strike price. The payoff will be done as long as the barrier has been hit.

55. Deferred Nominal Life Annuity

A monthly fixed-dollar payment beginning at retirement age. It is nominal because the payment is fixed in dollar amount at any particular time, up to and including retirement.

56. Deferred Payment Option

An option where the price paid is deferred until the end of the option's life.

57. Deferred Rebate Option

A claim that pays \$1 at expiration if the price of the underlying asset has reached a barrier prior to expiration. If a contract that pays \$1 at the time of a barrier is reached, it is called rebate option. Therefore, the deferred rebate option is similar to rebate option.

58. Deferred-Strike Options

Deferred-strike options are also called shout options. As the phrase "deferred-strike" implies, a shout option is an option whose strike price can be specified as an underlying asset price at any time before the maturity of the option. The level of the strike is ultimately set at a specific relationship to the spot, for example, 6% or 4% below the spot, or 4% or 6% above, during a period of time normally starting on the

trade date and ending on a date agreed upon at the trade time. After the strike is specified according to the terms in the contract or after the shouting time, the shout option becomes a vanilla option until the maturity of the option.

Shout options possess characteristics of American options. Since optimal timing or the “shouting” time is uncertain, there is no straightforward way to price shout options. However, they can be priced using either the binomial tree method or some analytical approximations.

59. Deferred Swap

A swap with terms specified today, but for which swap payments begin at a later date than for an ordinary swap.

60. Deferred Taxes

It is a noncash expense item. It results from differences between accounting income and true taxable income.

61. Deferred Up Rebate Option

A deferred rebate option for which the current stock price is below the rebate barrier. The payoff to this claim does not depend upon a strike price. The payoff will be done as long as the barrier has been hit.

62. Deficit

The amount by which a sum of money is less than the required amount; an excess of liabilities over assets, of losses over profits, or of expenditure over time.

63. Deficit-Budget Unit

An individual, business firm or unit of government whose current expenditures exceed its current receipts of income, forcing it to become a net borrower of funds in the money and capital markets.

64. Defined Benefit Plans

Pension plans in which retirement benefits are set according to a fixed formula. This plan promises in advance to pay employees a special level of benefit. A major question in the management and regulation of this kind of plan is whether an employer's contribution to the fund is sufficient to meet future pension liability.

65. Defined Contribution Plans

Pension plans in which the employer is committed to making contributions according to a fixed formula. However, benefits paid during retirement are not promised in advance. Instead they depend on contributions and earnings accumulated over time.

66. Degree of Combined Leverage (DCL)

A firm's degree of combined leverage (DCL) is the percentage change in earnings per share (EPS) that results from a 1% change in sales volume:

$$\text{DCL} = \frac{\text{Percentage change in EPS}}{\text{Percentage change in sales}}$$

The relationship between the degrees of operating and financial leverage and the degree of combined leverage is a multiplicative one. DFL times degree of operating leverage (DOL) results in:

$$\begin{aligned} & \frac{\text{Percentage change in EPS}}{\text{Percentage change in EBIT}} \times \frac{\text{Percentage change in EBIT}}{\text{Percentage change in sales}} \\ &= \frac{\text{Percentage change in EPS}}{\text{Percentage change in sales}} \end{aligned}$$

Thus, a firm's degree of combined leverage is simply the product of its degree of operating leverage and its degree of financial leverage. The DCL represents the impact on earnings per share of the combined effects of operating leverage and financial leverage if profit margins remain constant.

Like the degree of operating leverage and the degree of financial leverage, the DCL is not constant, as the firm's sales rise and fall over time. We know DOL declines as sales increase and DFL declines as EBIT rises. Thus, a firm's degree of combined leverage will fall as its sales and EBIT increase, as long as the firm's margin, fixed operating costs and financial costs remain constant.

The degree of combined leverage uses a firm's operating leverage and financial leverage and the assumption of constant margins to estimate a relationship between changes in sales and changes in earnings.

67. Degree of Financial Leverage (DFL)

A firm's financial risk reflects its interest expense, or in financial jargon, its financial leverage. A quick way to determine a firm's exposure to financial risk is to compute its degree of financial leverage. The degree of financial leverage (DFL) measures the sensitivity of EPS to changes in EBIT:

$$DFL = \frac{\text{Percentage change in EPS}}{\text{Percentage change in EBIT}}$$

This definition clearly suggests that DFL represents the percentage change in earnings per share due to a 1% change in earnings before interest and taxes.

There is a more straightforward method to compute a firm's degree of financial leverage that avoids handling percentage changes in variables. This formula is given as:

$$DFL = \frac{EBIT}{EBIT - I} = \frac{EBIT}{EBT}$$

DFL equals the firm's earnings before interest and taxes (EBIT) divided by EBIT minus interest expense (I), or earnings before taxes (EBT).

DFL changes with the level of EBIT for much the same reason that **DOL** changes with sales volume. [See also **Degree of operating leverage**] When EBIT is about the same as the firm's interest expense, EPS is small. A slight change in EBIT can therefore lead to a large percentage change in EPS, resulting in a large DFL. If the firm's interest expense does not change while EBIT continues to grow, the percentage increase in EPS becomes smaller and smaller, resulting in lower values for the firm's degree of financial leverage.

68. Degree of Operating Leverage (DOL)

A quick way to approximate a firm's exposure to business risk is to compute its degree of operating leverage. The degree of operating leverage (DOL) is a measure of the sensitivity of EBIT to a change in unit volume in sales, *assuming a constant price-variable cost margin*. Formally,

$$DOL = \frac{\text{Sales revenue} - \text{Variable costs}}{\text{Sales revenue} - \text{Variable costs} - \text{Fixed Costs}}$$

Or, stated with variables, the formula for DOL can be written as:

$$\frac{Q(p - v)}{EBIT} = \frac{Q(p - v)}{Q(p - v) - F}$$

where:

Q = quantity of goods sold

p = price per unit

v = variable cost per unit

F = total fixed cost

$EBIT$ = earnings before interest and taxes

These formulas make computing the degree of operating leverage seem fairly straightforward, but these calculations assume constant margins. Any careful analysis of business risk should include analysis of competitive conditions and other influences on the firm's margins.

Why does a firm's DOL change as its level of unit sales varies? Recall the basic definition of DOL: it is the percentage change in EBIT that corresponds to a 1% change in unit sales. For lower levels of sales, EBIT is small. At the firm's **break-even point**, EBIT is zero. [See also **Break-even point**] Therefore, any change in sales from the break-even point results in an infinite percentage change in EBIT and a DOL value of infinity. As sales volume grows, the level of EBIT also grows, but the resulting percentage change in EBIT becomes smaller and smaller, leading to reductions in the firm's degree of operating leverage. With constant margins and fixed costs, this implies that firm growth causes business risk to decline.

69. Delinquent Account

An account that is past due because the account holder has not made the obligated payment on time.

70. Deliverable Instrument

The asset in a forward contract that will be delivered in the future at an agreed-upon price. A forward contract for foreign exchange currency is a deliverable instrument.

71. Delivery

The tender and receipt of an actual commodity or financial instrument, or cash in settlement of a futures contract.

72. Delivery Date

Specific day that a futures contract expires.

73. Delivery Point

A point designated by a futures exchange at which the financial instrument or commodity covered by futures contract may be delivered in fulfillment of such contract.

74. Delivery Price

Price agreed to (possibly some time in the past) in a forward contract.

75. Delta

The change in the price of a derivative due to a change in the price of the underlying asset. Based upon the call option formula defined in option pricing model. [See also **Option pricing** equation] The mathematic result can be defined as:

$$\frac{\partial C}{\partial S} = N(d_1) > 0$$

76. Delta-Hedging

Hedging a derivative position using the underlying asset, with the amount of the underlying asset determined by the derivative's sensitivity (delta) to the price of the underlying asset.

77. Delta-Gamma Approximation

A formula using the delta and gamma to approximate the change in the derivative price due to a change in the price of the underlying asset.

78. Delta Neutral Portfolio

The value of the zero-delta portfolio is not affected by changes in the value of the asset on which the options are written.

79. Demand Deposit

Transactions account, payable on demand, that pays no interest to the depositor.

80. Demand Loan

A borrowing of funds (usually by a security dealer) subject to recall of those funds on demand by lender.

81. Demand Shock

An event that affects the demand for goods and services in the economy.

82. Denomination

Face value or principal of a bond.

83. Deposit Insurance

Deposit insurance is when a company pays an insurance fee to the treasury department or federal reserve bank. Deposit insurance is one of the significant

benefits of having an account at an FDIC-insured bank – it is how the FDIC protects your money in the unlikely event of a bank failure. The standard insurance amount is \$250,000 per depositor, per insured bank, for each account ownership category. And you do not have to purchase deposit insurance. If you open a deposit account in an FDIC-insured bank, you are automatically covered.

84. Deposit Multiplier

Also called the coefficient of deposit expansion because it indicates how many dollars of new deposits will result from an injection of one more dollar of excess reserves into the banking system.

85. Depository Institutions

Financial institutions that raise loanable funds by selling deposits to the public.

86. Depository Institutions Deregulation and Monetary Control Act (DIDMCA)

Law passed in 1980 by the US Congress to deregulate interest rate ceilings on deposits and grant new services to nonbank thrift institutions as well as to impose common reserve requirements on all depository institutions.

87. Depository Transfer Check (DTC)

A depository transfer check (DTC) is an ordinary check restricted “for deposit only” at a designated bank. Hence, the designated collection bank deposits a DTC for the daily deposits into the firm’s checking account and then submits the DTC to the collection system. Although the DTC is less expensive than a wire transfer, it is also slower.

88. Depreciation

A non-cash expense, such as the cost of plant or equipment, charge against earnings to write off the cost of an asset during its estimated useful life. It can use straight line method to do the depreciation. [See **Double-declining balance depreciation**]

89. Depreciation Tax Shield

The term $T(\text{Dep})$, the tax rate multiplied by the depreciation expense, is called the depreciation tax shield. It represents the tax savings the firm receives from its noncash depreciation expense. For example, with a 34% tax rate, a depreciation expense of \$1,000 reduces a firm’s tax bill by \$340.

90. Deregulation

The lifting or liberalization of government rules that restrict what private businesses (especially those in the financial markets) can do to serve their customers.

91. Derivative

A financial instrument whose value is determined by the specific features of the underlying asset or instrument. [See **Primitive security**]

92. Derivative Asset/Contingent Claim

Securities providing payoffs that depend on or are contingent on the values of other assets such as commodity prices, bond and stock prices, or market index values. Examples are futures and options.

93. Derivative Security

[See **Primitive security**]

94. Detachable Warrant

A warrant entitles the holder to buy a given number of shares of stock at a stipulated price. A detachable warrant is one that may be sold separately from the package it may have originally been issued with (usually a bond).

95. Development Projects

Development projects are attempts to develop projects and technologies that represent small advances of an already established knowledge base. These “sure things” will be low-risk, low-return investments in R&D.

96. DI System

The directional indicator system (DI system) is from a technical family known as momentum oscillators. Whereas the previous systems outlined deal with the futures-price level, oscillators deal with price changes. The logic employed by the directional- indicator system is that any trending period can be characterized as having a significant excess of either positive or negative price movements. Periods when prices are quickly moving upwards will have more upwards price change than downward price change, and vice versa. It is this relative price change that the DI estimates.

97. Diagonal Spread

A position in two calls where both the strike prices and times to maturity are different. It can be regarded as a combination of bull (or bear) and calendar spread. (A diagonal spread can also be created with put options.)

98. Differential Equation

An equation relating a variable to its derivatives and one or more independent variables. The differential equation can be classified into deterministic and stochastic differential equation. Black and Scholes (1973, pp. 637–659) have used stochastic differential equation to derive the option pricing model.

99. Diffusion Process

Generally, a continuous stochastic process in which uncertainty increases with time. Also used to describe the Brownian (random) part of an Itô process. [See **Differential equation**]

100. Digital Option

Another name for binary option. [See **Binary option**]

101. Diluted Common Shares Outstanding

The earnings statement of a firm reports both basic and diluted common shares outstanding. Diluted common shares outstanding include potential dilution of option-like securities. Therefore, diluted common shares outstanding are larger than basic common shares outstanding.

102. Dilution

Loss in existing shareholders' value. There are several kinds of dilution: (1) dilution of ownership, (2) dilution of market value, and (3) dilution of book value and earnings, as with warrants and convertible issues. Firms with significant amounts of warrants or convertible issues outstanding are required to report earnings on a "fully diluted" basis.

103. Direct Agency Costs

[See **Agency costs**]

104. Direct Finance

Any financial transaction in which a borrower and a lender of funds communicate directly and mutually agree on the terms of a loan.

105. Direct Lease

A lease under which a lessor buys equipment from a manufacturer and leases it to a lessee. In other words, it gives the lessee the use of an asset while the lessor retains title and ownership of the asset.

106. Direct Loan

Loan with terms negotiated directly between the lender and actual user of the funds.

107. Direct Paper

Short-term commercial notes issued directly to investors by borrowing companies without the aid of a broker or dealer.

108. Direct Placement

Sale of corporate securities to one or a small group of investors following private negotiations.

109. Direct Quote

A direct quote states an exchange rate in terms of the amount of US dollars that equal one unit of foreign currency, such as \$0.6786/DM. [See **Indirect quote**]

110. Direct Search Market

Buyers and sellers seek each other direct and transact directly.

111. Dirty Price

The present value of a bond's future cash flows (this implicitly includes accrued interest). For instance, a 7% annual coupon bond trading at par would have a dirty price of \$107 just prior to coupon payment. Credit Metrics estimates dirty prices since the coupon is paid in non-default states but assumed not paid in default. [See **CreditMetrics model** and **Accrued interest**]

112. Disbursing Float

A decrease in book cash but no immediate change in bank cash, generated by checks written by the firm. [See **Float**]

113. Discount

The forward rate either will be at a discount or a premium to the spot rate. A currency is selling at a discount if it can be purchased more cheaply in the forward than in the spot market. Or, in other words, using indirect quotes, a dollar that is selling at a discount can buy fewer units of the foreign currency in the forward market than in the spot market.

114. Discount Bonds

A bond that sells below par value is said to be selling at a *discount* and is called a discount bond.

The price of a discount bond will rise as it nears maturity if the market rate remains the same, since at maturity its price will equal its par value.

115. Discount Broker

A brokerage firm that offers a limited range of retail services and charges lower fees than full-service brokers.

116. Discount Factor

The term $1/(1 + r)^n$ is called a *present value interest factor*, $PVIF(r, n)$, or discount factor.

When the number of periods n and interest rate r are the same, the future value interest factor (FVIF) and PVIF terms are merely reciprocals of each other. That is,

$$PVIF = 1/FVIF \text{ and } FVIF = 1/PVIF$$

As it is the reciprocal of the compounding future value interest factor, the present value interest factor, $1/(1 + r)^n$, will diminish as either the interest rate or the number of years increases. Thus, using the same discount rate, cash flows in the far future are worth less to us than nearer cash flows. Over the same time frame, higher discount rates result in lower PVIFs, meaning future cash flows will be worth less in present value terms.

117. Discount Function

The discounted value of \$1 as a function of the time until payment.

118. Discount Instrument

An instrument, such as a Treasury bill, that provides no coupons.

119. Discount Method

A method of calculating the interest charge on a loan that deducts the interest owed from the face amount of the loan, with the borrower receiving only the net proceeds after interest is deducted for his or her use.

120. Discount Payback Period Rule

An investment decision rule in which the cash flows are discounted at an interest rate and the payback rule is applied on these discounted cash flows. This method has

taken time value of money into account. However, it still does not consider all potential cash flow.

121. Discount Rate

The interest charge (in annual percentage terms) set by the Federal Reserve banks for borrowings by depository institutions from the discount windows of these Reserve banks.

There are two possible meanings for this term as follows:

1. Occasionally, the Fed implements monetary policy by adjusting the discount rate, the interest rate it charges on its loans to banks. This serves to encourage or discourage banks from borrowing from the Fed to raise loanable reserve. Changes in the discount rate also transmit signals regarding future Fed policy.
2. The interest rate that is used to find a present value often is called a discount rate.

122. Discount Window

Interest rate charged by Federal Reserve banks lending to member institutions.

123. Discounted Cash-Flow Valuation Theory

It is the basic tool for determining the theoretical price of a corporate security. The price of a corporate security is equal to the present value of future benefits of ownership. For example, for common stock, these benefits include dividends received while the stock is owned plus capital gains earned during the ownership period.

124. Discounted Dividend Model (DDM)

A formula to estimate the intrinsic value of a firm by figuring the present value of all expected future dividends. [See **Gordon model**]

125. Discounting

Discounting is the process of determining the present value, or the value as of today, of a future cash flow.

126. Discretionary Account

An account of a customer who gives a broker the authority to make buy and sell decisions on the customer's behalf.

127. Disintermediation

The withdrawal of funds from a financial intermediary by the ultimate lenders (savers) and the lending of those funds directly to ultimate borrowers.

128. Disposition Effect

The disposition effect refers to investors' reluctance to sell assets that have lost value and greater likelihood of selling assets that have made gains. This phenomenon can be explained by prospect theory (loss aversion), regret avoidance and mental accounting.

129. Distressing Exchange

During a time of credit distress, debt holders may be effectively forced to accept securities in exchange for their debt claim—such securities being of a lower value than the nominal present value of their original claim. They may have a lower coupon, delayed sinking funds, and/or lengthened maturity. For historical estimation of default probabilities, this would count as a default event since it can significantly impair value.

130. Distribution

A type of dividend paid by a firm to its owners from sources other than current or accumulated retained earnings.

131. Diversifiable Risk

A risk that specifically affects a single asset or a small group of assets. [See **Unsystematic risk**]

132. Diversification

Diversification occurs when we invest in several different assets rather than just a single one.

Financial theorists commonly assume that the goal of a business is to maximize shareholder wealth. Hence decisions should be evaluated on the basis of how they affect value and, more directly, how they affect the amount and uncertainty of the cash flow stream accruing to the owners.

One line of financial theory has sought to explain conglomerate mergers through the diversification effect. The basic argument follows from portfolio theory: joining together two less than perfectly correlated income streams reduces the relative variability of the streams. However, it has been conversely argued that a perfect

capital market allows no economic advantage from a purely conglomerate merger. That is, a merger cannot create diversification opportunities beyond those available to an individual invest before the merger.

133. Diversification

Diversification in a portfolio includes stocks from different industries. Unsystematic risk will be diversified away.

134. Diversification Benefits

If a company merges with another company that produces a different product, this merge will enjoy diversification benefits.

135. Divestitures

In a divestiture, one firm sells a segment of its operations to another firm. No new corporate entity is created. The selling firm gives up the operational cash flows associated with the divested assets in exchange for a cash flow from the buyer. Arguably, a decision to divest a segment should be made in a capital budgeting framework.

Tax treatment of a divestiture handles the transaction as an ordinary sale with capital gains or losses recognized normally.

136. Dividend

A payment made to holders of a firm's common stock and/or preferred stock. Cash dividends are paid in cash while stock dividends are paid in stock.

137. Dividend Declaration Date

On the dividend declaration date, the directors of the firm may issue a statement declaring a regular dividend. The statement might be worded something like, "On January 2, 2005, the directors of this corporation met and declared quarterly dividends to be \$0.75 per share payable to the holder of record on January 22; payment will be made on February 7, 2005." With this declaration, the dividend becomes a legally binding obligation to the corporation.

138. Dividend Growth Model

A model wherein dividends are assumed to be at a constant rate in perpetuity. [See **Gorden model**]

139. Dividend Irrelevance

Miller and Modigliani (1961, pp. 411–432) were the first to present an argument for dividend irrelevance. Miller and Modigliani's theory that the value of the firm is independent of its dividend policy is similar to their analysis of the irrelevance of capital structure. The theory assumes a world without taxes or transaction costs. In addition, investors are assumed to be rational, with homogeneous expectations, and both corporate management and shareholders are assumed to know the same information about the firm.

140. Dividend Payout Ratio

It equals dividend per share divided by earnings per share. [See **Retention rate**]

141. Dividend Policy

Dividend policy is the decision a firm makes to pay out earnings or retain them for reinvestment in the firm. If it pays out dividends, company policy must determine the amount to retain. Two questions drive a firm's dividend policy: Does dividend policy have an effect upon the firm's value? If so, will the firm try to achieve an optimal payout ratio by attaining an ideal dollar payment per share? These questions have sparked debate between practitioners and academicians for many years. Practitioners see an optimal level of dividend payout, whereas some academic factions have argued that dividend policy does not affect the value of the firm at all.

Still other groups of academics have argued that dividends are the *only* factor that determines firm value. This shows up in Gordon's constant dividend growth model for a share of common stock:

$$P_0 = \frac{D_1}{r - g}$$

where:

P_0 = current stock price

D_1 = dividend payout in next period

r = cost of equity capital for the firm

g = growth rate for the firm

According to the Gordon model, if the firm increases its cash dividend, the price of its stock will increase. Remember, however, that any increase in the dividend is a reduction in retained earnings, which causes lower growth rate, g , for the firm. According to the model, a lower growth rate reduces the firm's stock price, so the optimal dividend policy must balance the effects of these two variables to maximize the stock price.

142. Dividends Per Share

Amount of cash paid to shareholders expressed as dollar per share.

143. Dividend Yield

Dividends per share of common stock divided by market price per share.

144. DMAC System

The dual moving-average crossover system (DMAC system) employs logic similar to the MAPB system by seeking to find when the short-term trend rises above or below the long-term trend. The MAPB represents the short-term trend by the daily price and the long-term trend by the moving average. The DMAC uses a short-term moving average and long-term moving average to represent the short- and long-term trend. A change in the price trend is signaled when these two moving averages cross. Specifically, a buy signal is generated when the shorter moving average is greater than (above) the longer moving average, and a sell signal when the shorter moving average is less than (below) the longer moving average. The trader always maintains a long or short position in the futures market.

145. Dollar-Weighted Return

The internal rate of return on an investment.

146. Dominance Principle

Under the efficient-frontier analysis, the assumption that an investor prefers returns and dislikes risks. For example, an individual is prepared to experience risk associate with the same risk, he or she can obtain a higher expected return with two different return portfolios. Thus, the higher return portfolio dominates the lower return portfolio and would be preferred. Similarly, if an individual was satisfied with a return, he would select the less risky portfolio.

147. DONCH System

The Donchian system (DONCH system) is part of a family of technical systems known as price channels. The system generates a buy signal any time the daily high price is outside (greater than) the highest price in the specified time interval. A sell signal is generated any time the daily high breaks outside (lower than) the lowest price in the same interval. The system always generates a signal for the trader to take a position, long or short, in the futures market.

148. Double-Declining Balance Depreciation

One of the accelerated depreciation methods. To use the double-declining-balance (DDB) method we first need to find the annual depreciation rate, which is calculated as follows:

$$\text{Annual depreciation rate} = \frac{1}{N}$$

where N is the number of years used to calculate straight line depreciation method. The annual straight-line rate of depreciation is $1/N$ percent per year; for DDB we need only multiply this amount by 2. Using this rate, the depreciation over 5 years for an asset with an initial value of \$6,000 is calculated as follows:

			Depreciation
Year 1	\$6,000 (2/5)	=	\$2,400
Year 2	(\$6,000 - \$24,000)(2/5)	=	1,440
Year 3	(\$6,000 - \$2,400 - \$1,440)(2/5)	=	864
Year 4	(\$6,000 - \$2,400 - \$1,440 - \$864)(2/5)	=	518
Year 5	(\$6,000 - \$2,400 - \$1,440 - \$864 - \$518)(2/5)	=	178 (311)
		Total	\$5,400

The maximum depreciation that can be taken is the value of cost minus salvage of \$5,400. Thus, \$178 of depreciation in year 5 exhausts the depreciation allowed under DDB, even though \$311 is available.

Comparison between straight-line and double-declining-balance method

Year	Straight-line	DDB
1	\$1,080	\$2,400
2	1,080	1,440
3	1,080	864
4	1,080	518
5	1,080	178

149. Double Taxation

Tax law complicates the dividend decision by imposing the burden of double taxation. In effect, income that a firm pays to shareholders as dividends is taxed twice. A corporate income tax is levied on the corporation's profits, and shareholders then pay personal income taxes on the dividends they receive. This is one of the complexities of the US tax laws that affect dividend policy.

The investor has no control over corporate tax effects. It is up to the corporate managers to reduce or defer tax payments as much as possible. However, the investor can influence the amount of personal taxes due on any dividend earnings.

Investors can reduce or defer taxes by buying low-dividend, high-growth stocks. Or, if they are tax exempt, investors can buy dividend-paying stocks.

Personal income taxes may affect an investor's preference for dividends or capital gains. When the tax rate on capital gains is substantially lower than the tax rate on personal income, then shareholders should prefer capital gains to dividends. This increases the firm's focus on retained earnings as equity financing. Investors still can realize homemade dividends through capital gains from sales of stock.

Different investors will have different preferences between dividends and capital gains. For example, average investors may prefer dividends because of the need for additional income, while wealthy investors may prefer capital gains because they currently do not need the income. Some large, tax-exempt institutions, such as pension funds, pay no taxes on their investment income, so they may be indifferent between dividends or capital gains. Other tax-exempt institutions, such as foundations and endowments, may favor current income to help meet budget needs.

Investors can defer receiving capital gains by holding stock; gains are received and become taxable only when stock is sold at a profit. Dividends offer less flexibility. Once the firm pays a dividend, the investor must pay taxes on this income. Therefore, the ability to defer taxes on capital gains may bias the investor against cash dividend payments.

150. Doubling Option

A sinking fund provision that may allow repurchase of twice the required number of bonds at the sinking fund call price. This is one aspect of sinking fund call different from conventional bond call.

151. Dow Jones Industrial Average Index (DJIA)

The DJIA is an arithmetic average of the stock prices that make up the index. The DJIA originally assumed a single share of each stock in the index, and the total of the stock prices was divided by the number of stocks that made up the index:

$$DJIA_t = \frac{\sum_{i=1}^{30} P_{it}}{\frac{30}{\sum_{i=1}^{30} P_{0i}}}$$

Today, the index is adjusted for stock splits and the issuance of stock dividends:

$$DJIA_t = \frac{\sum_{i=1}^{30} \frac{P_{it}}{AD_i}}{\sum_{i=1}^{30} P_{0i}}$$

Table 4 Adjustment of DJIA divisor to allow for a stock split

Before split		After 2-for-1 stock split by stock A
Stock	Price	Price
A	60	30
B	30	30
C	20	20
D	10	10
	120	90
Average before split = $\frac{120}{4} = 30$		
Adjusted divisor = $\frac{\text{sum of prices after the split}}{\text{Average before split}} = \frac{90}{30} = 3$		
Average after split = $\frac{90}{3} = 30$		
Before-split divisor = 4		After-split divisor = 3

where P_{it} = the closing price of stock i on day t , and AD_t = the adjusted divisor on day t . This index is similar to the simple price index except for the stock splits adjusted overtime. The adjustment process is illustrated in Table 4.

Alternatively, the average after split can be calculated as

$$\text{Average} = \frac{30 \times 2 + 30 + 20 + 10}{4} = 30$$

This average is identical to that obtained by using the adjusted-divisor approach. As Table 4 shows, the adjustment process is designed to keep the index value the same as it would have been if the split had not occurred. Similar adjustments have been made when it has been found necessary to replace one of the component stocks with the stock of another company, thus preserving the consistency and comparability of index values at different points in time.

152. Dow Theory

One of the tools used by technical analysts to measure supply and demand and forecast security prices is the Dow theory. The Dow theory is used to indicate reversals and trends in the market as a whole or in individual securities. According to the theory, there are three movements going on in the markets at all times. These movements are (1) daily fluctuations (the narrow movement from day to day), (2) secondary movements (short-run movements over 2 weeks to a month or more), and (3) primary trends, major movements covering at least 4 years in duration. The theory asserts that daily fluctuations are meaningless. However, daily asset prices or the market average must be plotted in order to outline the primary and secondary trends. In plotting the asset prices, the Dow theorists search for,” price patterns indicating market tops and bottoms.

Technical analysts use three basic types of charts: (1) line charts, (2) bar charts, and (3) point-and-figure charts. Bar charts have vertical bars representing each day’s price movement. Each bar spans the distance from the day’s highest price to the day’s lowest price with a small cross on the bar marking the closing price. Lines are used to

connect successive day's prices. Patterns indicating market tops or bottoms are then searched for in these line charts by technical analysis. *The Wall Street Journal* uses the bar charts to show daily fluctuations in the Dow Jones Average.

Point-and-figure charts are more complex than line or bar charts. These charts draw the percentage change directly. They are not only used to detect reversals in a trend but are also employed to set actual price forecasts.

The construction of a point-and-figure chart varies with the price level of the stock being charted. Only significant changes are posted to a point-and-figure chart. As a result there are one-point, two-point, three-point, and five-point point-and-figure charts.

To set the price target (forecasted stock price) which a stock is expected to attain, point-and-figure chartists begin by finding a congestion area. A congestion area is a horizontal band created by a series of reversals around a given price level. Congestion areas are supposed to result when supply and demand are equal. A breakout is said to have occurred when a column of price increase rises above the top of a congestion area. Breakout refers to a price rise or fall in which the price rises above or falls below the horizontal band which contained the congestion area. A penetration of the top of a congestion area is a signal for continued price rise. Penetration of the bottom of a congestion area by a column of price declines is a bearish signal.

To establish estimates of the new prices that a security should attain, point-and-figure chartists measure the horizontal width of a congestion area as they watch for a breakout. When a breakout occurs, the chartist projects the horizontal count upward or downward in the same direction as the breakout to establish the new price target.

153. Down-and-In Option

An option that comes into existence when the price of the underlying asset declines to a prespecified level. This option can be classified into down-and-in call and down-and-in put. For example, down-and-in call is a regular call that comes into existence only if the asset price reaches the barrier level.

154. Down-and-Out Option

An option that ceases to exist when the price of the underlying asset declines to a prespecified level. This option can be classified into down-and-out call and down-and-out put. For example, down-and-out call is a regular call that ceases to exist if the asset price reaches a certain barrier level.

155. Downgrade Trigger

A clause in the contract that states that the contract will be terminated with a cash settlement if the credit rating of one side falls below a certain level.

156. Downside Risk

In investment analysis, we calculate the mean and standard deviations of a company's rate of return. It is well known that we prefer companies with higher average returns and lower standard deviations. If the security has a positive rate of return distribution, then the downside risk is lower than that of a security with a negative distribution.

157. Draft

A written order requesting one party to make payment to another party at a specified point in time.

158. Drift

The expected change per unit time in an asset price.

159. Drift Rate

The average increase per unit of time in a stochastic variable. The drift can be undetectable amid all the up and down movements due to the random terms. [See **Stochastic process**]

160. Du Pont Analysis

Breaking return on equity into component parts is called Du Pont analysis. [See **Profitability ratios**]

161. Du Pont System of Financial Control

Highlights the fact that return on asset (ROA) can be expressed in terms of the profit margin and asset turnover. [See **Profitability ratios**]

162. Dual Banking System

Banking system in the United States in which groups trying to obtain a charter to open a bank can apply to the state banking department or the office of the Comptroller of the Currency-the national banking agency. Therefore, charters of US banks can be classified into state bank charter and national charter bank.

163. Dual Funds

Funds in which income and capital shares on a portfolio of stocks are sold separately.

164. Dumbbell Strategy

Dumbbell strategy are characterized by the inclusion of some proportion of short and intermediate term bonds that provide a liquidity buffer to protect a substantial investment in long-term security. The dumbbell portfolio divides its funds between two components. The shortest maturity is usually less than 3 years, and the longest maturities are more than 10 years. The portfolio is weighted at both ends at the maturity spectrum. The logic and mechanics of the dumbbell strategy are straightforward: the short-term treasury notes provide the least risk and highest liquidity, while long-term bonds provide the highest return. The best risk/return portfolio combination may very well be a combination of these extremes. Assuming an upward-sloping yield curve, no intermediate bonds will be held since they have (1) less return than the longest-maturity bonds, and (2) less liquidity and safety than the shortest T-note.

165. Duration

The weighted average time of an asset's cash flows. The weights are determined by present value factors. The formula can be defined as:

$$D = \frac{\sum_{t=1}^T \frac{(t)(PV_t)}{\sum_{t=1}^T PV_t} \quad (6)$$

where D is the duration of the bond; t is specific point in time; T is the number of years to maturity; PV_t is the present value of the cash flow received at time t . PV is further defined as follows:

$$PV_t = \frac{C_t}{(1+r)^t}, \quad (7)$$

where r is the interest rate, and C_t is the cash payment in period t ($t = 1, 2, 3, \dots, T$).

If coupon is zero, then duration is equal to maturity, therefore, the duration of zero-coupon bond is equal to maturity.

Use Eqs. 6 and 7 for an example shown in Table 5 to calculate duration.

For the 6% coupon bond, we find the duration in terms of Eq. 8 as:

$$\text{Duration} = (.0586)(1) + (.9414)(2) = 1.9414 \text{ years}$$

Similarly, the duration of the 12% coupon bond in Table 5 is 1.9047 years. The duration of 6% coupon bond is longer than that of 12% coupon bond where the maturity for two coupon bonds is equal. In sum, duration refers to the weighted average life of the bond, which also provides a measure of bond's sensitivity of interest rate changes. Two common duration measures are modified and Macaulay duration. The duration discussed here is the Macaulay duration which the yield curve is assumed to be flat.

Table 5 Calculation of duration for two bonds

6% Coupon bond				
Year	Payment	Present value (At 9.91%)	Proportion of present value	Proportion in year X
1	60	54.59	0.0586	0.0586
2	1,060	877.47	0.9414	1.8828
		932.06	1.0000	1.9414
12% Coupon bond				
Year	Payment	Present value (At 9.82%)	Proportion of present value	Proportion in year X
1	120	109.27	0.1053	0.1053
2	1,120	928.66	0.8947	1.7894
		1,037.93	1.0000	1.8947

166. Duration Gap (DUR_{GAP})

The weighted duration of assets (DUR_A) minus the product of the weighted duration of liabilities (DUR_L) and the ratio of total liabilities to total assets (W_L). The formula can be defined as:

$$DUR_{GAP} = DUR_A - (W_L)(DUR_L)$$

where W_L = total liability/total assets.

167. Duration Matching

A procedure for matching the durations of assets and liabilities in a financial institution.

168. Duration Measure

Duration measure is simply a weighted-average maturity, where the weights are stated in present value terms. In the same format as the weighted-average term to maturity, duration is

$$D = \frac{PVCF_1}{PVTCF}(1) + \frac{PVCF_2}{PVTCF}(2) + \dots + \frac{PVCF_n}{PVTCF}(n)$$

where:

$PVCF_t$ = the present value of the cash flow in year t discounted at current yield to maturity

t = the year when cash flow is received

n = maturity

$PVTCF$ = the present value of total cash flow from the bond discounted at current yield to maturity

169. Dyl Model

Dyl introduced short selling with margin requirements by creating a new set of risky securities, the ones sold short, which are negatively correlated with the existing set of risky securities. These new securities greatly enhance the diversification effect when they are placed in portfolios. The Dyl model affects the efficient frontier in two ways: (1) If the investor were to combine in equal weight any long position in a security or portfolio with a short position in a security or portfolio, the resulting portfolio would yield zero return and zero variance. (2) Any combination of unequal weighted long or short positions would yield portfolios with higher returns and lower risk levels.

170. Dynamic Asset Allocation

Dynamic asset allocation is a portfolio management strategy that frequently adjusts the mix of asset classes to suit market conditions. Adjustments usually involve reducing positions in the worst-performing asset classes while adding to positions in the best-performing assets.

171. Dynamic Capital Budgeting Decision

If a capital budgeting decision explicitly considers the change in interest rates, economic condition changes, and other environmental condition changes, then this is considered a dynamic capital budgeting decision.

172. Dynamic CAPM

Dynamic CAPM allows the for the investment opportunity set to change over time. For example, the consumption CAPM is a dynamic type of CAPM which allows investment opportunities to change over time.

173. Dynamic Conditional Correlation

The dynamic conditional correlation (DCC) model is a class of multivariate models. The model has the flexibility of univariate GARCH models coupled with parsimonious parametric models for the correlations. They are not linear but can often be estimated very simply with univariate or two step methods based on the likelihood function. It is shown that they perform well in a variety of situations and give sensible empirical results.

174. Dynamic Financial Ratio Analysis

In basic finance and accounting courses, industry-average ratios are usually used as a benchmark with which to compare a specific company's ratio at a specific point of time. This is a form of static ratio analysis because the focus is on one point in time.

But making static comparisons between ratios does not take full advantage of all the information the ratios provide. Dynamic analysis helps us better compare the ratios between either two firms or between the ratio of individual firm and that of industry average. In addition, this kind of relationship can be used to forecast the future ratios.

The financial manager compares the firm's ratios against same norm, such as the industry's average ratios. Let us take the debt ratio (DR) as an example. By regressing the current year's debt ratio against the industry average of debt ratio, the manager can better analyze the dynamic nature of ratios and determine the adjustment process the firm should undertake to get back on target. If the firm's DR is off target, the manager would attempt to adjust it to meet the mark. Lev (1969, pp. 290–299) developed the partial adjustment model to define the dynamic financial ratio adjustment process. The equation is

$$Y_{j,t} = Y_{i,t-1} + \delta_j Y_{j,t}^* - Y_{j,t-1} \quad (8)$$

where $Y_{j,t}^*$ = desirable target ratio for firm j ; $Y_{j,t-1}$ = previous period's ratio for firm j ; δ_j = partial adjustment coefficient for firm j reflecting technological and institutional constraints; and $Y_{j,t}$ = current year's ratio for the firm.

The partial adjustment model takes the difference between the firm's debt ratio and the target ratio (industry average) and adjusts it by δ . The difference can be only partially adjusted because deviations caused by financial and capacity constraints cannot be completely removed in the short run. Therefore, the coefficient of adjustment reflects the fact that there are limitations to the periodic adjustment of ratios.

175. Dynamic Hedging

A procedure for hedging an option position by periodically changing the position held in the underlying assets. The objective is usually to maintain a delta-neutral position. It can be quite expensive because of the transaction costs involved. Dynamic hedging is sometimes referred to as dynamic options replication. [See **Static option replication**]

176. Dynamic Option Replication

Option replication can be classified into either static or dynamic replication. Dynamic replication requires the position in the hedging assets to be rebalanced frequently and can be quite expensive because of the transaction costs involved. [See **Static option replication** and **Delta hedging**]

1.5 E

1. EAC Method

Equivalent annual cost method is used to do capital decision for projects with unequal lives. It equals to the NPV (N) of cost divided by an annuity factor that

has the same life as the investment. However, this method assumes that two alternative projects have same revenue. [For a general case see **Equivalent annual NPV**.]

2. EAFE Index

The European, Australian, Far East index, computed by Morgan Stanley, is a widely used index of non-US stocks.

3. Early Exercise

Exercise prior to the maturity date.

4. Early Withdrawal Penalty

An interest penalty a depositor pays for withdrawing funds from a deposit account prior to maturity.

5. Earning Assets

Income-earning assets held by a bank; typically include interest-bearing balances, investment securities, and loans.

6. Earnings Credit

The assumed interest rate at which a bank applied to customer's investable balances to earn interest income. Estimating investment income from balances involves four steps.

1. The bank determines the average *ledger (book) balances* in the account during the reporting period.
2. The average transactions float-uncollected funds that still appear as part of the customer's ledger deposit-is subtracted from the ledger amount. This difference equals *collected balances*.
3. The bank deducts required reserves that must be maintained against collected balances to arrive at investable balances.
4. Management applies an earnings credit rate against investable balances to determine the average interest revenue earned on the customer's account.

7. Earnings Dilution

A decrease in earnings per share after one bank acquires another.

8. Earnings Forecasts

In security analysis, earnings forecast is one of the most important actions to be taken. In general, we can use either a regression model or an ARIMA model. ARIMA is autoregression integrated with moving average model.

9. Earnings Management

Earnings management is the use of accounting techniques to produce financial statements that present an overly positive view of a company's business activities and financial position. Many accounting rules and principles require that a company's management make judgments in following these principles. Earnings management takes advantage of how accounting rules are applied and creates financial statements that inflate or "smooth" earnings. Earnings management included accrual earnings management and real earnings management.

10. Earnings Per Share

Net income divided by the number of outstanding shares of common stock.

11. Earnings Retention Ratio

It equals one minus pay-out ratio. [See **Plowback ratio**]

12. Earnings Revisions

An earnings-revisions investment strategy is based on empirical evidence that stocks with earnings forecasts that were recently upgraded by analysts tend to outperform, while stocks with negative earnings revisions tend to underperform.

13. Earnings Surprises

An earnings surprise occurs when a company's reported quarterly or annual profits are above or below analysts' expectations. These analysts, who work for a variety of financial firms and reporting agencies, base their expectations on a variety of sources, including previous quarterly or annual reports and current market conditions, as well as the company's own earnings' predictions or "guidance."

14. Earnings Yield

The ratio of earnings to price, E/P.

15. EBIT

Earnings before interest and taxes.

16. EBITDA

Earnings before interest, taxes, depreciation, and amortization.

17. EBIT/EPS Analysis

EBIT/EPS analysis allows managers to see how different capital structures affect the earnings and risk levels of their firms. Specifically, it shows the graphical relationship between a firm's operating earnings, or earnings before interest and taxes (EBIT) and its earnings per share (EPS). Scenario analysis with different levels of EBIT can help analysts to see the effects of different capital structures on the firm's earnings per share.

EBIT/EPS analysis is an older tool that was first developed when accounting concepts dominated financial analysis. Also, most managers are familiar with the concept of earnings and are more comfortable discussing the impact of leverage on earnings rather than on cash flow.

Leverage obviously affects earnings per share. For low values of EBIT, the proposed capital structure leads to lower EPS than the current structure. For higher values of EBIT, debt works to the firm's benefit, as EPS is higher under the proposed capital structure than under the current structure.

EBIT/EPS analysis has several practical implications. First, it shows the ranges of EBIT where a firm may prefer one capital structure over another. Second, should the expected EBIT of the firm lie above the indifference EBIT level, the firm's managers should examine the standard deviation of their EBIT forecast. If there is a relatively high probability that the actual EBIT level may fall below the indifference level, management may decide to play it safe and use a more conservative financing strategy with less debt.

Third, a firm's level of business risk will affect its desired exposure to financial risk. Variations in firm sales lead to changes in EPS through the joint effects of operating and financial leverage, as given by the following relationship:

$$\text{DOL} \times \text{DFL} = \text{DCL}$$

where DOL, DFL, and DCL represent degree of operating leverage, degree of financial leverage and degree of combined leverage, respectively.

A firm's DFL, or degree of financial leverage, is related to its choice of capital structure. Other things being equal, as a firm uses more debt to finance its assets, its degree of financial leverage rises.

There is no evidence that firms adjust their DOLs and DFLs to match some standard degree of combined leverage. This relationship does, however, imply a potential trade-off between a firm's business and financial risk. Firms with volatile sales, variable price-cost margins, and large amounts of fixed operating expenses may prefer to use less debt in their capital structures. A firm with excessive risk will not attract stock or bond investors.

EBIT/EPS is a simple tool, and so it has a limitation: it provides little insight into how financing decision affect shareholder wealth. Still, it provides managers with an idea of how different levels will affect earnings and earnings variability.

18. Econometric Model

There are two alternative ways to define econometric model:

- (a) It is based on representations of the underlying economic behavioral system for a particular commodity. These representations attempts to identify and model the relevant supply-and-demand factors that together determine market price and quantity.
- (b) The use of systems of equations and statistical estimation methods to explain or forecast changes in interest rates or other variables.

19. Economic Assumptions

Economic environment in which the firm expects to reside over the life of the financial plan. The economic condition can be classified as boom, normal, or recession.

20. The real flow of cash that a firm could pay out forever in the absence of any change in the firm's productive capacity.

21. Economic Income

Economic income is defined as the maximum value that a firm can consume in a given period and be as well off at the end of the period as it was at the beginning. **Accounting income** measures only the changes in wealth caused by realized or recognized gains and losses, revenues and expenses. [See also **Accounting income**] Economic income measures changes in wealth based upon both realized and unrealized gains and losses. This is why the market value of the firm (its stock price) usually differs from its book value.

Theoretically, financial analysis should consider economic income rather than accounting earnings to determine the value of a firm, since economic income represents the firm's true earnings and cash flows. However, economic income is not directly observable. For that reason, analysts generally use accounting earnings as a proxy. The relationship between economic income and accounting earnings can be related by the following equation:

$$\text{Accounting Income} = \text{Economic Income} + \text{Error}$$

22. Economic Value Added (EVA)

Economic value added (EVA) is a tool by which division managers can correct for failures of accounting-driven or sales-driven evaluation systems. EVA addresses the

shortcomings of these performance measures while at the same time including a cost most measures omit – the cost of capital, or the cost of financing the firm's operations with debt and equity. EVA is roughly equal to after-tax operating profit minus the firm's dollar cost of capital. If EVA is positive, management has added value to the firm; if it is negative, shareholder wealth has been harmed. In sum, EVA is a measure of financial performance trade-marked by Stern, Stewart & Co. equal to a firm's net operating profit after tax (NOPAT) minus a capital charge representing the required return to shareholders.

23. Economics, Relationship to Finance

The field of economics provides the basic framework within which managers make firm-level decision, since microeconomic decisions are implemented in the context of a dynamic, global macro economy. Like economics, finance employs the theory of rational decision making; like quantitative management science, finance does use some highly structured models and methods.

24. Economies of Scale and Economies of Scope

Economies of scale and high capital requirements typically go together. Scale economies occur as average production cost declines with rising output per period. Scale economies occur as average production cost declines with rising output per period. Any new entrant must (1) have available financing to construct a large-scale factory and (2) be able to sell in sufficient quantity to be cost-competitive. Entry may be especially unattractive when the entrant considers the impact of added volume on market price; the increase in supply caused by a new entry may lower product prices, making it more difficult for the new entrant to compete in the market. Scale requirements can deter entry and promote positive net present value projects among existing firms. Economies of scope, in particular, refers to financial institution's abilities to generate synergistic cost savings through joint use of inputs in producing multiple products.

25. ECU

European Currency Unit

26. ECU Swap

Used to transform principal and coupon payments denominated in European Currency Units into another currency, and vice versa.

27. Edge Acts

Special subsidiaries of US banking organizations authorized by federal law and regulation to offer international banking services.

28. Edge Act Corporation

A specialized organization form open to US domestic banks since 1919 and foreign banks since 1978. These banks specialized in international trade-related banking transactions or investments.

29. Effective Annual Interest Rate

A way of quoting an interest rate such that the quoted rate is the annual percentage increase in an amount invested at this rate. If \$1 is invested at an effective annual rate of r , the payoff in 1 year is $\$1 + r$. [See also **Effective annual rate**]

30. Effective Annual Rate (EAR)

The effective annual rate (EAR) sometimes called the *annual effective yield* is the true opportunity cost measure of the interest rate, as it considers the effects of periodic compounding. If the periodic interest charge r is known, the EAR is found by:

$$EAR = (1 + r)^m - 1,$$

where m = the number of compounding periods per year

If the **annual percentage rate (APR)** is known instead, r is found by dividing the APR by m and compounding by m periods. [See also **Annual percentage rate (APR)**]

$$EAR = \left[1 + \frac{APR}{m} \right]^m - 1$$

It is useful to distinguish between a contractual or stated interest rate and the group of rates we call yields, effective rates, or market rate. A contract rate, such as the annual percentage rate (APR), is an expression that is used to specify interest cash flows such as those in loans, mortgages, or bank savings accounts. The yield or effective rate, such as the effective annual rate (EAR), measures the opportunity costs; it is the true measure of the return or cost of a financial instrument.

31. Effective Annual Yield

Annualized interest rate on a security computed using compound interest techniques. [See also **Effective annual rate**]

32. Effective Convexity

The value for convexity that reflects the price impact of embedded options in different interest rate environments. [See **Convexity**]

33. Effective Duration

The value for duration reflecting the price impact of embedded options when interest rates rise versus fall. In addition, it can also refer to percentage change in bond price per change in the level of market interest rate.

34. Efficiency Ratio

Noninterest expense divided by the sum of net interest income and noninterest income. This is an aggregate profitability measure for a bank.

35. Efficient Diversification

The organizing principle of modern portfolio theory, which maintains that any risk-averse investor will search for the highest expected return for any level of portfolio risk.

36. Efficient Frontier

Graph representing a set of portfolios that either (a) maximize expected return at each level of portfolio risk or (b) minimize risk at each level of return.

37. Efficient Market

Unexpectedly good or bad news can cause assets' prices to change. Good news surprises lead market participants either to reduce the risk premium they demand of an asset (thus decreasing its required return) or to increase their expectations for future cash flows. Either reaction leads to an increase in an asset's price. Bad news surprises lead the market to demand a higher risk premium (and required return) or to reduce its expectations for future cash flows; either reaction results in a falling asset price.

If a market adjusts prices quickly and in an unbiased manner after the arrival of important news surprises, it is said to be an efficient market. If the market, for example, for IBM stock is efficient, we should see a quick price change shortly after any announcement of an unexpected event that affects sales, earnings, or new products. A quick movement in the price of a stock such as IBM should take no longer than several minutes. After this price adjustment, future price changes should appear to be random. That is, the initial price reaction to the news should be unbiased, or, on average, fully reflect the effects of the news. Every time IBM's stock price changes in reaction to new information, it should show no continuing tendency to rise or fall after the price adjustment.

Any consistent trend in the same direction as the price change would be evidence of an **inefficient market** that does not quickly and correctly process new information to properly determine asset prices. Likewise, evidence of price corrections or

reversals after the immediate reaction to news implies an inefficient market that overreacts to news.

In an efficient market, it is difficult to consistently find stocks whose prices do not fairly reflect the present values of future expected cash flows. Prices will change only when the arrival of new information indicates that an upward or downward revision in this present value is appropriate.

This means that in an efficient market, investors cannot consistently profit from trades made after new information arrives at the market. The price adjustment occurs so rapidly that no buy or sell order placed after the announcement can, in the long-run, result in returns above the market's average return. An order to buy after the arrival of good news may result in large profits, but such a gain will occur only by chance, as will comparable losses. Stock price trends always return to their random ways after initially adjusting to the new information.

Efficient markets result from interactions among many market participants, all analyzing available information in pursuit of an advantage. Also, the information flows or news they analyze must be random, both in timing and content (i.e., in an efficient market, no one can consistently predict tomorrow's news). The profit motive leads investors to try to buy low and sell high on the basis of new information and their interpretation of it. Hordes of investors analyzing all available information about the economy and individual firms quickly identify incorrectly priced stocks; resulting market pressures immediately push those stocks to their correct prices. In an efficient market, this causes prices to move in a *random walk*, meaning that they appear to fluctuate randomly over time, driven by the random arrival of new information.

38. Efficient Market Hypothesis (EMH)

The prices of securities fully reflect available information. Investors buying securities in an efficient market should expect to obtain an equilibrium rate of return. Weak-form EMH asserts that stock prices already reflect all information contained in the history of past prices. The semistrong-form hypothesis asserts that stock prices already reflect all past and current publicly available information. The strong-form hypothesis asserts that stock prices reflect all relevant information, including insider information.

39. Efficient Portfolio

If a portfolio is efficient, there exists no other portfolio having the same expected return at a lower variance of returns. Moreover, a portfolio is efficient if no other portfolio has a higher expected return at the same risk of returns.

40. Efficient Set

Graph representing a set of portfolios that maximize expected return at each level of portfolio risk. Each point on an efficient set represents an efficient portfolio.

41. Elasticity

A measure of the relative quantity response to a change in price, income, interest rate, or other variable.

42. Elasticity (of an Option)

Percentage change in the value of an option accompanying a 1% change in the value of a stock.

43. Electronic Trading

System of trading where a computer is used to match buyers and sellers.

44. Electronic Transfer

An electronic transfer is essentially a high-tech, automated **depository transfer check (DTC)**. [See also **Depository transfer check**] To speed the cash transfer, an electronic check image is processed through clearinghouses rather than through a wire network. An electronic transfer is cheaper than a DTC and usually clears in a single business day.

45. Embedding Option

An option that is an inseparable part of another instrument. For example, bond with embedded options. Such bonds are debt exchangeable for common stock (DECS), premium equity participating shares (PEPS) and preferred equity redeemable for common stock (PERCS). All of these instruments are effectively bonds plus some options position.

46. Empirical Research

Research based in historical market data.

47. Employee Stock Ownership Plans

Employee ownership can help align the incentives of *all* the firm's workers with those of the shareholders. Employee stock ownership plans (ESOPs), if correctly implemented, can make all employees think and act like owners. As part of the ESOP, employee participation groups (comprised of members elected by fellow employees) meet regularly to discuss ways to increase productivity and the firm's value and, therefore, the value of the employees' stake. About 10,000 firms, many of them privately owned, currently have ESOPs in place.

ESOPs can be difficult to implement for an international firm. Some countries ban stock options or limit ownership of foreign shares. For a firm residing in such a country, an ESOP plan needs to be designed to meet specific, local requirements.

48. EMU

European Monetary Union. There are 11 euro-zone countries. In this union, the European Central bank sets the monetary policy.

49. End-of-Year Convention

Treating cash flows as if they occur at the end of a year (or, alternatively, at the end of a period), as opposed to the date convention. Under the end-of-year convention, the end of year 0 is the present, end of year 1 occurs one period hence, and so on.

50. Endowment Funds

Organization chartered to invest money for specific purposes.

51. Enhancement

Enhancement is less common than **cannibalization**; it reflects an increase in the cash flows of the firm's other products that occurs because of a new project. [See also **Cannibalization**] For example, adding a delicatessen to a grocery store may increase cash flows more than the deli sales alone if new deli customers also purchase grocery items.

52. Enterprise Value

The value of a firm equal to the market capitalization (market value of the equity) plus the market value of outstanding debt.

53. Enterprise Value – Earnings Before Interest Tax and Amortization (EBITDA) Ratio

Enterprise value is an estimate of the market value of the company's operating assets, which means all the assets of the firm except cash. Since market values are usually unavailable, we use the right-hand side of the balance sheet and calculate the enterprise value. Enterprise value is equal to Total Market Value of Equity plus Book Value of Total Liabilities minus Cash. Notice that the sum of the value of the market values of the stock and all liabilities equals the value of the firm's assets from the balance sheet identity. Total market value of equity equals market price per share

times basic number of shares outstanding. EBITDA ratio equals Enterprise value divided by EBITDA. EBITDA is equal to Net Income plus Interest Expense plus Tax expense plus depreciation and amortization. This ratio is similar to the PE ratio, but it relates the value of all the operating assets to a measure of the operating cash flow generated by those assets.

54. Equal Credit Opportunity Act

A federal law passed in 1974 forbidding lending institutions from discriminating in the granting of credit based on the age, race, ethnic origin, religion, or receipt of public assistance of the borrowing customer.

55. Equilibrium Model

A model for the behavior of interest rates derived from a model of the economy.

56. Equilibrium Rate of Interest

The interest rate that clears the market.

57. Equipment Obligation Bonds

[See **Collateral**]

58. Equipment Trust Certificate

An equipment trust certificate gives the bondholder a claim to specific “rolling stock” (moveable assets), such as railroad cars or airplanes. The serial numbers of the specific items of rolling stock are listed in the bond indenture and the collateral is periodically examined by the trustee to ensure its proper maintenance and repair.

59. Equity

Ownership interest of common and preferred stockholders in a corporation. Also, total assets minus total liabilities, or net worth.

60. Equity-Like Signal

As a general rule, companies prefer to issue equity when they think their shares are overvalued. And investors prefer to purchase equity when they are confident that the asset in question will create value for the buyer, since the seller will have a stake in the buyer after the sale. Thus, a buyer’s offer to pay with equity sends a negative signal about the buyer, but a seller’s acceptance of equity signals the seller’s positive private information about the asset.

61. Equity-Residual Method

The equity-residual method is formulated in a manner that emphasizes the goal of financial management, which is the pursuance of the interests of shareholders. It can be interpreted as stating that the benefit of the project to the shareholders is the present value of the cash flows not going to pay operating expenses or to service or repay debt obligations. With these flows identified as those going to shareholders, it is appropriate, and rather easy, to discount these flows at the cost of equity. The only difficulty involved is identifying this cost of equity, a problem embodied in all capital-budgeting methods. See Appendix O for further detail.

62. Equity Kicker

Used to refer to warrants because they usually are issued in combination with privately placed bonds.

63. Equity Method

One of the two methods that accounted for stock held as an investment in another corporation. The equity method is used if the investing firm exercises significant control over the other corporation (investee). Under this method the investment is recorded at cost. Any net earnings of the investee are recorded in proportion to the investor's share of ownership as an increase in the investment account of the investor. Dividends or net losses of the investee result in a decrease in the investing firm's investment account.

64. Equity Multiplier

Calculated as assets divided by total equity; the equity multiplier is determined by the firm's financing policy. A firm that uses a larger amount of financial leverage can support a faster sustainable growth rate, when all else remains constant. If actual growth exceeds the sustainable growth rate, a firm can finance the difference by taking on additional debt. Growth below the planned rate may lead to smaller additions to debt and an unplanned reduction in financial leverage. [See also **Capital structure ratios**]

65. Equity Swap

A swap where the return on an equity portfolio is exchanged for either a fixed or a floating rate of interest.

66. Equity-Linked Forward

A forward contract (e.g., for currency) where the quantity to be bought or sold depends upon the performance of a stock or stock index.

67. Equivalent Annual NPV (EANPV)

The net present value of a project divided by an annuity factor that has the same life as the investment.

We will give an example to demonstrate how different lives in capital budgeting needs an annuity factor to deal with this issue. In addition we will mathematically derive the annuity factor. The traditional NPV technique is suitable for investment projects that have the same life. However, it may not be appropriate to select a project from mutually exclusive investment projects, if these projects have different lives. The underlying reason is that, compared with a long-life project, a short-life project can be replicated more quickly in the long run. In order to compare projects with different lives, we can compute the NPV of an infinite replication of the investment project. For example, let Projects A and B be two mutually exclusive investment projects with the following cash flows.

Year	Project A	Project B
0	-100	-100
1	70	50
2	70	50
3		50

By assuming a discount rate of 12%, the traditional NPV of Project A is 18.30 and the NPV of Project B is 20.09. This shows that Project B is a better choice than Project A. However, the NPV with infinite replications for Projects A and B should be adjusted into a comparable basis.

In order to compare Projects A and B, we compute the NPV of an infinite stream of constant scale replications. Let $NPV(N, \infty)$ be the NPV of an N -year project with $NPV(N)$, replicated forever. This is exactly the same as an annuity paid at the beginning of the first period and at the end of every N years from that time on. The NPV of the annuity is:

$$NPV(N, \infty) = NPV(N) + \frac{NPV(N)}{(1+K)^N} + \frac{NPV(N)}{(1+K)^{2N}} + \dots \quad (9)$$

In order to obtain a closed-form formula, let $(1/[(1+K)^N]) = H$. Then we have:

$$NPV(N, t) = NPV(N)(1 + H + H^2 + \dots + H^t). \quad (10)$$

Multiplying both sides by H , this becomes

$$H[NPV(N, t)] = NPV(N)(H + H^2 + \dots + H^t + H^{t+1}). \quad (11)$$

Subtracting Eq. 11 from Eq. 10 gives:

$$NPV(N, t) - (H)NPV(N, t) = NPV(N)(1 - H^{t+1}),$$

$$NPV(N, t) = \frac{NPV(N)(1 - H^{t+1})}{1 - H}.$$

And taking the limit as the number of replications, t , approaches infinity gives:

$$\begin{aligned} \lim_{t \rightarrow \infty} NPV(N, t) &= NPV(N, \infty) \\ &= NPV \left[\frac{1}{1 - [1/(1 + K)^N]} \right] \\ &= NPV(N) \left[\frac{(1 + K)^N}{(1 + K)^N - 1} \right] \end{aligned} \quad (12)$$

Equation 12 is the NPV of an N -year project replicated at constant scale an infinite number of times. We can use it to compare projects with different lives because when their cash-flow streams are replicated forever, it is as if they had the same (infinite) life.

Using Eq. 12 for Projects A and B gives us the following:

For Project A

$$\begin{aligned} NPV(2, \infty) &= NPV(2) \left[\frac{(1 + 0.12)^2}{(1 + 0.12)^2 - 1} \right] \\ &= (18.30) \left[\frac{1.2544}{0.2544} \right] \\ &= 90.23. \end{aligned}$$

For Project B

$$\begin{aligned} NPV(3, \infty) &= NPV(3) \left[\frac{(1 + 0.12)^3}{(1 + 0.12)^3 - 1} \right] \\ &= 20.09 \left[\frac{1.4049}{0.4049} \right] \\ &= 69.71. \end{aligned}$$

Consequently, we would choose to accept Project A over Project B, because, when the cash flows are adjusted for different lives, A provides the greater cash flow.

$$K \times NPV(N, \infty) = \frac{NPV(N)}{\text{Annuityfactor}}, \quad (13)$$

Alternatively, Eq. 12 can be rewritten as an annuity version as: where the annuity factor is

$$\left[1 - 1/(1 + K)^N\right]/K.$$

The decision rule from Eq. 13 is equivalent to the decision rule of Eq. 12. The left-hand side of Eq. 13 is defined as equivalent annual NPV, which is called the equivalent annual NPV method in capital budgeting decision.

68. Equivalent Loan

The amount of the loan that makes leasing equivalent to buying with debt financing in terms of debt capacity reduction. This concept can be used to determine whether a firm should buy or lease equipment.

69. Equivalent Taxable Yield

The pretax yield on a taxable bond (t) providing an after-tax yield equal to the rate on a tax-exempt municipal bond (r_m).

$$r(1 - t) = r_m$$

$$r = r_m/(1 - t)$$

where t = marginal tax rate

Thus the equivalent taxable yield (r) is simply the tax-free rate (r_m) divided by $1 - t$.

70. ERISA

Employee Retirement Income Security Act of 1974. This is a federal law that governs the administration of pension plans for nongovernmental employees. The basic provision is that all private corporations fully fund their pension plans.

71. Erosion

Cash-flow amount transferred to a new project from customers and sales of other products of the firm.

72. Estimation Risk

The risk of error in estimating a project's cash flows or required rate of return is called forecasting risk or estimation risk. The following table reviews some source of estimation risk associated with cash flows and required rate return.

Expected cash flows	Required rates of return
Political risk	Real risk-free return
Blocked currencies	Supply/demand for funds
Tariffs, quotas, embargoes	Macroeconomic consumption patterns

(continued)

Expected cash flows	Required rates of return
Military conflict	Investor optimism/pessimism
Unstable government	Long-run real economic growth
Fluctuating exchange rates	
Central bank policy	Expected inflation
	Monetary policy
Fiscal policy	Commodity prices
Government spending	
Tax policy	Risk premium
	Systematic risk
Inadequate or incorrect	Political risk
Strategic analysis	Exchange rate risk
Market research	Business risk
Pricing policy	Financial risk
Competitor retaliation	
Construction delays	
Delay in R&D, manufacturing, or production	
Work stoppages or strikes	
Technology obsolescence	

73. Euro

The European currency unit introduced in January 1999.

74. Eurobanks

Banks that make loans and accept deposits in foreign currencies.

75. Eurobonds

Eurobonds are bonds denominated in US dollars issued by firms in financial markets outside the United States and typically pay interest annually. Eurobonds are an international bond sold primarily in countries other than the country in whose currency the issue is denominated.

76. Eurocurrency

A currency that is outside the formal control of the issuing country's monetary authorities.

77. Eurocurrency Deposits

Deposits of funds in a bank that is denominated in a currency foreign to the bank's home country.

78. Eurocurrency Loans

Loans made by a multinational bank in a currency other than that of the bank's home country.

79. Eurocurrency Market

An international money market where bank deposits denominated in the world's most convertible currencies are traded.

80. Eurodollars

Dollar-denominated deposits at banks of located outside the United States. Eurodollar transaction denotes any transaction involving dollars that takes place outside the United States.

81. Eurodollar Bonds

Eurodollar bonds are dollar-denominated bonds that are underwritten by international syndicates of commercial and investment banks. Because these issues are sold outside the United States, they escape review by the SEC, somewhat reducing their issue expenses. Eurodollar bonds usually have fixed coupons with annual coupon payments. Most mature in 3–10 years, so they are not attractive for firms that want to issue long-term debt. They typically are unsecured, pledging no specific assets to the bondholders in case of default. This is not a major concern to investors, as only the largest and financially strongest firms have access of the Eurobond market. Investors do care that the bonds are sold in bearer form, thus helping bondholders to remain anonymous and evade taxes on coupon income. Some researchers believe that this is the main reason that Eurodollar bond interest rates are low relative to US rates.

82. Eurodollar CD

Deposit of dollars with foreign banks.

83. Eurodollar Futures Contract

A futures contract written on a Eurodollar deposit.

84. Eurodollar Interest Rate

The interest rate on a Eurodollar deposit.

85. Euroequity

Firms are not limited to domestic financial markets for raising capital. The world's financial markets have become more liquid and more integrated as cross-

border restrictions have diminished, and more and more large corporations have begun looking outside their national boundaries to raise financing. US firms can raise money in the Euroequity market by selling equity and debt claims to non-US investors. Changes in tax laws and regulations, as well as lower financing costs, have led US firms to issue more offerings to non-US investors.

A cost advantage of trading in the Euroequity market is that Euroequity is traded over the counter in a large, active, cross-border market, so issuing firms need not register their securities on exchanges in many different countries.

86. European, Australian, Far East (EAFE) Index

A widely used index of non-US stocks computed by Morgan Stanley.

87. European Currency Unit (ECU)

An index of foreign exchange which was introduced in 11 European countries in January 2002.

88. European Monetary Unit (EMU)

It is a form of denomination. It is consisted of the currencies of six original European Currency Unit (EEC) members.

89. European Option

A European option is an option that can be exercised only on the expiration date, which makes it simpler to analyze as compared to an **American option** because its term to maturity is known. An American option may be exercised any time up to the expiration date. [See also **American option**] The factors that determine the values of American and European options are the same; all other things being equal, however, an American option is worth more than a European option because of the extra flexibility it grants the option holder.

90. Event Study

Research methodology designed to measure the impact of an event of interest on stock returns.

91. Ex-Dividend Date

A practical problem arises if a shareholder decides to sell a day or two before the record date. Because the brokerage industry requires some time to process the transaction and enter the name of the buyer on the stockholder list, it has arbitrarily decided that the right to the declared dividend is terminated four business days before the record date. Any sale between this ex-dividend date and the record date leaves the seller with the right to the dividend. The term ex-dividend comes from the

Latin *ex* meaning *from*, because the dividend has been taken from anyone who buys the stock after the ex-dividend date.

The extent of the appropriate drop in stock prices associated with dividend payments depends in part on the tax situation of the marginal investor, the individual who at the margin causes an imbalance between supply and demand and therefore causes a price change. If the marginal investor in the marketplace is in the 30% tax bracket and the dividend is \$1.00 per share, the per-share price of the stock might fall by \$0.70 on the ex-dividend date, all else being equal. The price might not fall by a full dollar because the marginal investor realizes only a \$0.70 after-tax dividend.

One model has been developed to incorporate tax effects into determining the ex-dividend price:

$$\frac{P_o - P_x}{D_o} = \frac{1 - T_p}{1 - T_g}$$

where:

P_o = the price just before the stock goes ex

P_x = the ex-dividend share price

D_o = the amount of the dividend per share

T_p = the relevant marginal personal tax rate

T_g = the effective marginal tax rate on capital gains

If $T_p = T_g = 0$ or $T_p = T_g$ then $P_x = P_o - D_o$

Tax laws require the corporation to mail a copy of Form 1099 to every shareholder at the end of the year to report the amount of dividends the firm paid to that person. The firm also sends a copy of this form to the IRS to report the dividend income it paid to each shareholder during the year. This system of informing the taxing authorities is unique to the United States. Most other nations of the world require that corporations withhold portions of stockholders' dividends and turn these funds over to the government to settle each individual's tax liability on dividend income.

In sum, ex-dividend date is a date four business days before the date of record for a security. An individual purchasing stock before its ex-dividend date will receive the current dividend.

92. Ex Rights or Ex Dividend

Phrases used to indicate that a stock is selling without a recently declared right or dividend. The ex-rights or ex-dividend date is generally four business days before the date of record.

93. Excess Reserves

Cash and deposits at the Federal Reserve banks held by depository institutions that are in excess of their legal reserve requirements.

94. Excess Return

Rate of return in excess of the risk-free rate.

95. Exchange Offering

A method used in the past by the US Treasury to sell new securities by offering them to investors who hold maturing Treasury securities.

96. Exchange Option

An option permitting the holder to obtain one asset by giving up another. For example, an exchange call maturing t period from today provides the right to obtain one unit of the Nikkei index in exchange for one unit of the S&P index. In addition, standard calls and puts are exchange options in which one of the two assets is cash.

97. Exchange Rate

An exchange rate is the price of one currency in terms of another currency. [See **Direct** or **Indirect quotes**]

98. Exchange Rate Risk

Investors in nondomestic securities face a number of risks beyond those of domestic securities. Exchange rate changes will cause fluctuations in the values of cash flows in terms of US dollars; this is called exchange rate risk.

99. Exchange Ratio for Business Combination

In business combination, two companies agree to exchange shares of common stock. In such a case, the determination of a “price” is actually the determination of an exchange ratio. Larson and Gonedes (1969, pp. 720–728) have presented a model for exchange ratio determination that involves making assumptions about the pre- and postcombination earnings streams and P/E (price/earnings) ratios.

For example, assume Firm A and Firm B are the acquired firm and the acquiring firm, respectively. Let the exchange ratio (ER) be defined as the number of traded shares of Firm B to be exchanged for the one traded share of Firm A. LG defined the postcombination price (p^*) as:

$$p^* = \left[\frac{NI_A + NI_B}{N_B + (ER)(N_A)} \right] (PE^*),$$

where:

NI_A = net income for Firm A

NI_B = net income for Firm B

N_A = number of shares outstanding for Firm A

N_B = number of shares outstanding for Firm B

PE^* = post combination price/earnings ratio

By comparing p^* with price per share before the combination of Firm A (P_A) and Firm B (P_B), we find the following exchange ratio will affect the shareholders of both Firm A and Firm B as follows:

(a) The shareholders of Firm A are as well off after the combination as before if:

$$ER \geq \frac{N_B P_B}{(N_A + N_B)(PE^*) - N_A P_A}$$

(b) The shareholders of Firm B are as well off after the combination as before if:

$$ER \leq \frac{(PE^*)(N_A + N_B)}{(N_A)(P_B)} - \frac{N_B}{N_A}$$

100. Exchanges

National or regional auction markets providing a facility for members to trade securities. A seat is a membership on an exchange. For example, New York exchange and Philadelphia exchange are national and regional exchange, respectively.

101. Exclusionary Self-Tender

The firm makes a tender offer for a given amount of its own stock while excluding targeted stockholders. It is the opposite of a targeted repurchase.

102. Executive Stock Options

Executive stock options provide stock purchase rights as compensation for corporate employees. For services rendered, the manager or the employee has the right to buy a specific number of shares for a set price during a given period. Unlike warrants and publicly traded options, executive stock options cannot be traded. The option's owner has only two choices: exercise the option or let it expire. Like a warrant, should the owner decide to exercise the option, the corporation receives money and issues new shares.

The use of executive stock options for management compensation raises an interesting agency question. The firm's managers may make investment and financing decisions that increase the firm's risk in order to increase the value of their stock options. Such an action could have a detrimental effect on the bondholders and other creditors of the firm.

103. Executor

An individual or trust department responsible for handling a settlement.

104. Exercise

The exchange of the strike price (or strike asset) for the underlying asset at the terms specified in the option contract.

105. Exercise Price

Price at which the holder of an option can buy (in the case of a call option) or sell (in the case if a put option) the underlying stock. Also called the striking price.

106. Exercise Style

The circumstances under which an option holder has the right to exercise an option. “European” and “American” are exercise styles.

107. Exercising the Option

The act of buying or selling the underlying asset via the option contract.

108. Exotic Option

A derivatives contract in which an ordinary derivative has been altered to change the characteristics of the derivative in a meaningful way. Also called a *nonstandard option*. Most exotic options trade in the over-the-counter market and designed by financial institutions to meet the requirement of their clients. For example, barrier options are exotic options. [See **Barrier option**]

109. Expectations Hypothesis

The expectations hypothesis assumes that bond investors look ahead and make predictions, or form expectations, about future interest rates. From this perspective, in an efficient market, the return from investing in an N -year bond will be the same as the expected return from rolling over the proceeds (coupons and principal) from maturing 1-year bonds into new 1-year bonds over the N -year time frame. Thus today's long-term rates reflect expectations about future short-term rates.

Although intuitive, the expectations hypothesis does not totally explain the shapes of observed term structures. Historically, the term structure is sloped upward; long-term rates usually are higher than short-term rates. Under the expectations hypothesis the typical upward-sloping term structure implies that the market always expects rising short-term interest rates. This does not agree with the

observed behavior of short-term rates over time. Other explanations for the behavior of the term structure have attempted to correct this flaw.

110. Expected Return

Risk arises from the possibility that actual returns may differ from expected returns. Actual returns differ from expected returns whenever there is an unexpected change in an asset's price or cash flow stream. Issuers must compensate investors with an expected return that is greater than the nominal risk-free return; otherwise, investors would have no economic incentive to place their capital at risk. A basic principle of finance is that higher risk leads to higher expected returns, or that *risk drives returns*.

A more complete model of returns is:

$$\text{Expected return} = (1 + \text{Real risk-free rate}) \times (1 + \text{Expected inflation rate}) \times (1 + \text{Risk premium}) - 1$$

Combining with our nominal risk-free rate gives us:

$$\text{Expected return} = (1 + \text{Nominal risk-free rate}) \times (1 + \text{Risk premium}) - 1$$

The term *expected return* on the left-hand side of the equations indicates that the investor may not earn the stated return on an instrument. Because of risk, the actual return may be higher or lower than expected.

111. Expected Return-Beta Relationship

One implication of the capital asset pricing model (CAPM) is that security risk premiums (expected excess returns) will be proportional to beta. This is used to describe relationship between return and systematic risk, as shown in the security market line (SML).

112. Expected Value of a Variable

The average value of the variable obtained by weighting the alternative values by their probabilities.

113. Expected Yield

The weighted average return on a risky security composed of all possible yields from the security multiplied by the probability that each possible yield will occur.

114. Expiration Date

The time by which the option transaction must be carried out. In other words, it is the maturity date of an option.

115. Expiration-Date Risk

Futures contracts are not usually available for every month. If a hedger needed a futures contract for July and the only contracts that were available were for March, June, September, and December, the hedger would have to select either the June or September contract. Either of these contract would have a different price series than a July contract (if one existed). Hence, the hedger cannot form a perfect hedge and is faced with the chance that the basis may change.

116. Explicit Finite Difference Method

A method for valuing a derivative by solving the underlying differential equation. The value of the derivative at time t is related to three values at time $t + \delta t$. It is essentially the same as the trinomial tree method.

117. Exposure

The amount which would be lost in a default given the worst possible assumptions about recovery in the liquidation or bankruptcy of an obligor. For a loan or used facility it is the full amount of the facility, since the worst assumption is that the borrower draws the full amount and then goes bankrupt. In a credit risk analysis, it is called exposure at default (EAD).

118. Extendable Notes

Notes that have their coupons reset every 2 or 3 years to reflect the current interest rate environment and any changes in the firm's credit quality. At each reset, the investor may accept the new coupon rate (and thus effectively extend the maturity of the investment) or put the bonds back to the firm.

119. Extendable Swap

In an extendable swap, one party has the option to extend the life of the swap beyond the specified period.

120. Extension

Voluntary arrangements to restructure a firm's debt, under which the payment date is postponed.

121. Extension Risk

The risk that the holder of a mortgage-backed security will receive outstanding principal payment later than originally anticipated. Later principal payments result from interest rates rising and prepayments occurring slower than expected.

122. External Financing

External financing is any kind of business funding you acquire from sources outside the company. Bank loans, investments from private individuals or investment firms, grants, selling company shares, and issue corporate bonds are all examples of external financing.

123. Extinguish

Retire or pay off debt.

1.6 F

1. Face Value

The value of a bond that appears on its face. Also referred to as par value or principal.

2. Facility

A generic term which includes loans, commitments, lines, letter, etc. Any arrangement by which a bank accepts credit exposure to an obligor.

3. Facility Fee

Fee imposed for making a line of credit available.

4. Factor

A financial institution that buys a firm's accounts receivables and collects the debt. [See **Factoring**]

5. Factoring

Firms can convert accounts receivable to cash by a method called factoring. Factoring essentially involve an outright sale of accounts receivable to a finance company or factoring department of a commercial bank. Factoring differs from **pledging** since it gives the finance company no recourse to the borrower in the case of bad debts. [See also **Pledging**] The customer receives notice that the invoice has been sold and is asked to make payment directly to the finance company.

This arrangement clearly increases the lender's risk, as compared to pledging. To reduce this risk, the finance company virtually takes over the work of the borrower's credit department. All new customer orders pass through the finance company, which does a credit appraisal. If the finance company rejects the customer as an

unacceptable credit risk, the borrower either must turn down the order or fill it for cash.

Factoring, like pledging, is a fairly costly source of credit. This overall cost has a number of distinct components. The factor charges the borrower a fee between about 1% and 3% of the face value of the invoices for credit appraisal. The interest charge depends upon whether the finance company has agreed to forward the funds as soon as the goods are shipped or only on the receivable's due date at the end of the credit period. For payment at shipment, the interest rate may well rise as high as 15–25%.

The advantages of factoring resemble those of pledging; it is a relatively easy and flexible source of funds once the initial negotiations have been completed, and it provides additional funds as the borrower's scale of operations, and therefore its needs, grow. Factoring always has been widely used by small companies in specific industries, such as textiles, garments, or furniture, which may lack access to bank loans. Factoring allows the smaller company to avoid the cost and trouble of setting up its own credit department; this gives factoring one advantage over pledging. Against this, however, must be set the possible damage to the borrower's reputation when customers learn that their accounts have been sold to a finance company.

Credit-card transactions share some common traits with factoring. In effect, the merchant that accepts a credit card payment is factoring its accounts receivable to the issuer of the card, and the credit-card holder pays the issuer directly.

In sum, factoring is a sale of a firm's accounts receivable to a financial institution known as a factor.

6. Factor Analysis

An analysis aimed at finding a small number of factors that describe most of the variation in a large number of correlated variables. (Similar to a principal components analysis.) [See **Arbitrage pricing theory**]

7. Factor Model

A model in which each stock's return is generated by common factors, called the systematic sources of risk. [See **Arbitrage pricing model**]

8. Factor Portfolio

A well-diversified portfolio constructed to have a beta of 1.0 on one factor and a beta of zero on any other factor. Factor portfolios will serve as the benchmark portfolios for a multifactor security market line.

9. Fair Credit Billing Act

A federal law giving customers the right to question entries on bills sent to them for goods and services purchased on credit and giving them the right to expect that billing errors will be corrected as quickly as possible.

10. Fair Credit Reporting Act

A federal law that gives credit customers the right to view their credit record held by a credit bureau and to secure quick correction of any errors in that record.

11. Fair Game

Under the fair game, the expected value of a gamble is exactly equal to the cost. Under this situation, there is no way to use “information” available at a point in time (t) to earn return above normal.

12. Fair Game Model

Bases on average returns across a large number of observations, the expected return on an asset equals its actual return that is

$$z_{j,t+1} = r_{j,t+1} - E(r_{j,t+1} | \Phi_t)$$

and $E(z_{j,t+1}) = E(r_{j,t+1} - E(r_{j,t+1} | \Phi_t)) = 0$ in which $z_{j,t+1}$ is the error term between the j th stock's actual return $r_{j,t+1}$ at time $t+1$ and its expected return $E(r_{j,t+1} | \Phi_t)$. The fair-game model is an expected return efficient-market model. In search of a fair game, investors can invest in securities at their current prices and can be confident that these prices fully reflect all available information and are consist with the risks involved.

13. Fair Market Value

Amount at which common stock would change hands between a willing buyer and a willing seller, both having knowledge of the relevant facts; also called market price.

14. Fair Value

Another name for the theoretical forward price; spot price plus interest less the future value of dividends.

15. Fallen Angels

Obligors having both relatively high percentage risk and relatively large exposure, whose large exposures were created when their credit ratings were better, but who now have much higher percentage risk due to recent downgrades.

16. Fama and French Factor Models

The Fama and French Three-Factor Model (or the Fama French Model for short) is an asset pricing model developed in 1992 that expands on the capital asset pricing

model (CAPM) by adding size risk and value risk factors to the market risk factor in CAPM. This model considers the fact that value and small-cap stocks outperform markets on a regular basis. By including these two additional factors, the model adjusts for this outperforming tendency, which is thought to make it a better tool for evaluating manager performance. In addition, Fama and French also developed a four factor and a five factor model.

17. Fannie Mae

Name referring to the Federal National Mortgage Association (FNMA).

Originally created in 1983, the Federal National Mortgage Association (FNMA), or “Fannie Mae,” is the oldest of the three mortgage-backed security sponsoring agencies. While it is now a private corporation owned by shareholders with stock traded on major exchanges, in the minds of many investors it still has implicit government backing that makes it equivalent to a government-sponsored agency. Indeed, supporting this view is the fact that FNMA has a secured line of credit available from the US Treasury should it need funds in an emergency. FNMA is a more active agency than Government National Mortgage Association (GNMA) in creating pass-through securities. GNMA merely sponsors such programs. FNMA actually helps create pass-throughs by buying and holding mortgages on its balance sheet; it also issues bonds directly to finance those purchases.

18. FASB Statement 13

Complicated accounting rules must guide a presentation of the effects of a financial lease on the lessee’s balance sheet and income statement. FASB Statement 13 requires the firm to capitalize lease payments and list the leased property as an asset (“leased property under capital lease”) and as a liability (“obligations under capital lease”). The rationale for this accounting treatment is straightforward. Since the financing lease is a long-term, fixed obligation to the firm, it should be treated like similar liabilities. The firm’s long-term access to the leased asset and liability accounts are amortized to zero over time under the FASB 13 treatment. The income statement deduction includes both the amortization of the liability and an imputed interest expense on the remaining lease liability.

19. Feasible Set

Opportunity set in a portfolio analysis.

20. Fed Wire

The Federal Reserve System operates by Fed Wire to process US-dollar-denominated transactions initiated and received in the United States. Since Fed Wire transfers are guaranteed by the US government, the system minimizes users’ liquidity and credit risks.

21. Federal Agencies

Departments or divisional units of the federal government empowered to borrow funds in the open market in order to make loans to private businesses and individuals or otherwise subsidize private lending or borrowing.

22. Federal Agency Securities

Securities issued by corporations and agencies created by the US government, such as the Federal Home Loan Bank Board and Government National Mortgage Association (Ginnie Mae).

23. Federal Deposit Insurance Corporation (FDIC)

The FDIC was created in the Glass-Steagall Act (1933). Fed member bank must be insured by the FDIC. If a banks purchases deposit insurance, it must comply with rules set by FDIC.

24. Federal Financing Bank (FEB)

FEB is a federal agency that borrows from the US Treasury and lends funds to various federal agencies. FEB can require the treasury to purchase up to \$5 billion of its obligations. The treasury secretary is authorized to purchase any amount of FEB obligations at his or her discretion.

25. Federal Funds

Unsecured short-term loans that are settled in immediately available funds. Federal funds are excess reserves lent by one institution to another institution to meet Fed reserve requirements.

26. Federal National Mortgage Association (FNMA)

A federal agency created in 1938 to buy and sell selected residential mortgages in the secondary market and thus encourage the development of a resale market for home loans.

27. Federal Open Market Committee (FOMC)

The chief body for setting and monitoring money and credit policy within the Federal Reserve System, consisting of the seven members of the Federal Reserve Board and the presidents of the 12 Federal Reserve banks, only five of whom may vote.

28. Federal Reserve Bank

One of the 12 district federal reserve banks that make up the Federal Reserve System. Typically, regional Federal Reserve banks serve as clearing locations where institutions accounts are debited or credited as necessary, and checks are sorted, bundled, and returned to participating depositories.

29. Federal Reserve Board

The Federal Reserve Board (or the Fed) acts as the central bank, or “banker’s bank,” in the US economy.

The Federal Reserve Board has three basic instruments with which it can affect the money supply to administer its **monetary policy**. [See also **Monetary policy**]

1. **Open market operations** (repurchases or sales of government securities)
2. **Discount rate** changes (adjustment in the interest rate paid by banks when they borrow from the Fed)
3. **Reserve requirement** changes (adjustments in the amount of reserves banks must hold either as cash or on deposit at the Fed)

The only interest rate the Fed directly controls is the discount rate. As with any price, demand and supply influences affect all other interest rates. [See also **Open market operations**, **Discount rate**, and **Reserve requirement**]

30. Federal Reserve Statement

A weekly listing of the factors supplying reserves to and absorbing the reserves of depository institutions.

31. Federal Reserve System

The central bank of the United States, created by Congress to issue currency and coin, regulates the banking system, and takes measures to protect the value of the dollar and promote full employment.

32. Federal Savings and Loan Insurance Corporation (FSLIC)

The federal agency charged with insuring the deposits accepted by member savings and loan associations up to a maximum of \$100,000 for each account.

33. Feedforward Neural Networks

A feedforward neural network is an artificial neural network wherein connections between the nodes do not form a cycle. As such, it is different from its descendant:

recurrent neural networks. The feedforward neural network was the first and simplest type of artificial neural network devised.

34. Feltham–Ohlson Model

Feltham and Ohlson demonstrated analytically, using dynamic information, that losses, particularly at the stage of start-up in growth and technology-based companies, are considered to be costs that create an effect of conservatism accounting, consequently, there is an undervaluation of assets, hence the results and equity. However, this situation tends to be reversed over time, because given the principle of rationality, the investors continue to invest in the company if those investments are associated with abnormal profitability expectations.

35. FHA

Federal Housing Administration- a federal agency that insures mortgages which target groups that might otherwise be disadvantaged in the housing market, such as low income family.

36. FHLMC

Federal Home Loan Mortgage Corporation (Freddie Mac); a private corporation operating with an implicit federal guarantee; buys mortgages financed largely by mortgage-backed securities. The implicit Federal guarantee can reduce the borrowing cost.

37. Fiber Optics

Electronic cables capable of transmitting data and images over great distances using bursts of laser light.

38. Fidelity Bond

A contract that covers losses associated with employee dishonesty, typically embezzlement and forgery at banks.

39. Fiduciary

An individual or trust department responsible for acting in the best interests of a designated third party.

40. Field Warehouse Financing

A form of inventory loan in which a public warehouse company acts as a control agent to supervise the inventory for the lender.

41. Field Warehousing

In field warehousing, the finance company (usually a specialized warehousing organization) takes over the use of a certain part of the borrower's premises. This floor space must be segregated from the borrower's other operations so that it can be kept locked, restricting access only to the warehousing company. The inventory to serve as collateral is transferred to this segregated area, and the warehousing company advances the discounted cash value of the inventory to the borrower. In return, the warehousing company receives a warehouse receipt, which gives it title to the inventory.

This inventory cannot be sold or used without the warehouse company's permission, and this permission is given only when the borrower repays a corresponding portion of the funds advanced. Thus, the lender can ensure that the collateral always is adequate to secure the loan. The warehousing company locates a member of its own staff, the custodian, on the borrower's premises to ensure that its rights are respected.

42. FIFO

The first-in first-out accounting method of inventory valuation. In an inflation period, the cost of inventory is lower than that calculated by the last-in-first-out accounting method.

43. Filter Rule

A technical analysis technique stated as a rule for buying or selling stock according to past price movements. The filter rule is usually stated in the following way: Purchase the stock when it rises by $X\%$ from the previous low and hold it until it declines by $Y\%$ from the subsequent high. At this point, sell the stock short or hold cash.

Filter rules are a timing strategy. They show investors when they should be long in a security and when they should sell it short. The alternative to timing is to buy and hold the security. Thus, filter rules are analyzed by comparing them to buy and hold strategy. One further assumption is necessary for the buy and hold strategy to be relevant, namely, the expected return is positive. If the expected return is negative, then the relevant alternative is to hold cash.

44. Finance

Finance is the study of how to manage assets and obtain funds in order to maximize the wealth of the owner. Thus, the broad field of finance deals with such varied topics as designing a personal retirement plan, managing inventory, investing excess cash, borrowing money, or attracting bank depositors. Business operations generate profits when the firm can raise funds at a lower cost than the return generated by the investment of the funds.

Businesses purchase assets with the hope that they will generate future cash flows. The cash flows may be in the form of income, future cost savings, and/or changes in company value. To finance asset purchases, firms sell liability and equity securities, including bonds, stocks, mortgages, and loans. Investors are willing to buy the securities in order to receive future cash flows, which help the investors meet their own future needs. The value of an asset depends on three cash flow characteristics: (1) amount; (2) pattern over time; and (3) risk.

Investors will pay more for an asset that promises larger cash flows after shorter time periods with lower risks. Values are lower for assets that generate smaller, later, and/or more uncertain cash flows.

45. Finance Charge

As defined by truth-in-lending Regulation Z the finance charge refers to “all charges payable directly or indirectly by the borrower and imposed directly or indirectly by the lender as an incident to or as an extension of credit.”

46. Finance Company

A firm that borrows from the money and capital markets to make loans to individuals and commercial enterprises. The services provided by finance companies include consumer lending, business lending, and mortgage lending. Finance companies do not accept deposits but instead rely on short- and long-term debt as a source of fund. Additionally, finance companies often lend money to customers who commercial banks find too risky

47. Financial Accounting Standards Board (FASB)

The governing body in accounting. FASB issues generally accepted accounting principles (GAAP) as a guide for financial statement reporting in the United States.

48. Financial Analyst

The position of the financial analyst in the corporate structure and the scope of his or her work are interdependent. The financial analyst is a staff member who diagnoses the effects of management proposals and/or decisions on the financial health of the firm. Acting as an internal consultant, the financial analyst examines profitability, cash flows, and operations; conducts studies; interprets information; and designs financial controls. Although some of this analysis is focused entirely within the firm, the analyst also must examine the dynamic economic, social, political, and competitive environments that are external to the firm, in an attempt to gauge their impact on the firm's well-being. This information is used to assist in the process of financial planning and forecasting. The analyst provides this information as input for upper-level management's decisions; generally, he or she does

not set policy or make decisions. Major decisions are made by top management, which may include the CFO and treasurer of the firm.

In addition, the financial analyst must perform many tasks on a periodic basis. These activities include analyzing the company's liquidity and profitability and supervising its day-to-day financial operations, including accounts receivable, accounts payable, and cash balances. The analyst must also contribute to longer term projects by analyzing the firm's capital structure and major investment alternatives.

The analyst also completes specific projects that are either self-initiated or, more commonly, requested by others. For example, if the analyst notices a market variation from a normal financial ratio, he or she may try to determine the underlying cause of the variation and report it to management as part of the control function. Also, the analyst may examine the effect of a current economic force on the company, such as how a tax policy change might affect the firm's cash flows and stock value. The financial analyst may assist operations management in determining whether to lease or purchase a specific asset. Some problem analyses are critical to the success of the entire company, such as the decision whether to expand or sell off one of the operating division; others are as commonplace as deciding whether to purchase Treasury bills or certificates of deposit (CDs) with surplus cash.

Within the company's organization structure, the position of financial analyst may be centralized or decentralized or have elements of both. Centralizing the analyst function places it at corporate headquarters, separate from the operational units for which it performs most of its analyses. Decentralizing the position places analysts in each of the firm's divisions to do division-specific work. Centralization allows the firm to pool expertise, promote interaction among the analysts, and maintain objectivity, as the analysis views a divisional issue from a companywide point of view consistent with the firm's overall strategy.

However, certain circumstances create advantages for a decentralized financial analyst function. A decentralized organization is useful: (1) when the analyst's role is to advise the operating manager, who has some independence to make decision-level decisions; (2) when the operations of the division are complex and the analyst must possess specialized expertise to make useful recommendations; and (3) when the larger firm is really a holding company for different, independent organization (e.g., one corporation may operate a banking division, another may run an insurance division).

49. Financial Assets

Financial assets such as stocks and bonds are claims to the income generated by real assets or claims on income from the government.

50. Financial Break-Even

Financial break-even occurs when the project breaks even on a financial basis, that is, when it has a net present value of zero. To determine a project's financial

break-even point, we must first determine the annual operating cash flow, OCF^* , that gives it a zero NPV. The formula for the formula for the financial break-even quantity:

$$Q_{\text{financial}}^* = \frac{FC + OCF^*}{p - vc} = Q_{\text{cash}}^* + \frac{OCF^*}{p - vc}$$

where:

FC = Fixed costs

VC = Variable cost per unit

P = Price per unit

OCF = Annual operating cash flow

Q_{cash}^* = cash break-even point

Without any calculations, we know intuitively that this break-even quantity should exceed the cash and accounting break-even quantities. OCF^* must be sufficiently large to both cover depreciation expense (Dep) and allow the project to earn its minimum required return. Intuition also tells us that accounting income under financial break-even should exceed that of the accounting break-even point. As OCF^* must exceed the depreciation expense, the firm's net income (NI) will be positive (ignoring working capital effects, $OCF = NI + \text{Dep}$). Thus, some positive taxable income occurs under financial break-even.

As expected the financial break-even quantity and operating cash flow exceed those of the **cash** and **accounting break-even** analyses. [See also **Cash break-even** and **Accounting break-even**] Note a major difference between financial break-even as compared to cash and accounting break-even: Financial break-even analysis encompasses cash flows from the *entire life* of the project.

51. Financial Constraints

For the investor, a financial constraint is any factor that restricts the amount or quality of investment options. They can be internal or external. Every investor faces both internal and external constraints. Some constraints are common sense. Each investor needs to understand his or her own time-horizon constraints, for example.

52. Financial Crisis

In a financial crisis, asset prices see a steep decline in value, businesses and consumers are unable to pay their debts, and financial institutions experience liquidity shortages. A financial crisis is often associated with a panic or a bank run during which investors sell off assets or withdraw money from savings accounts because they fear that the value of those assets will drop if they remain in a financial

institution. Other situations that may be labeled a financial crisis include the bursting of a speculative financial bubble, a stock market crash, a sovereign default, or a currency crisis. A financial crisis may be limited to banks or spread throughout a single economy, the economy of a region, or economies worldwide. A financial crisis may have multiple causes. Generally, a crisis can occur if institutions or assets are overvalued, and can be exacerbated by irrational or herd-like investor behavior. For example, a rapid string of selloffs can result in lower asset prices, prompting individuals to dump assets or make huge savings withdrawals when a bank failure is rumored. For example, in 2007 and 2008 the US and other financial markets experienced a severe financial crisis.

53. Financial Disclosure

The provision of relevant financial information to the public to aid individuals and institutions in making sound financial decisions.

54. Financial Distress

Financial distress means that a firm's short-run operating and financial cash inflows are less than its outflows.

Financial distress occurs when the firm's internal rate of return on its investments is less than its cost of capital – either at the present time or in the near future. In terms of sources and uses of funds, financial distress occurs when the inflow of funds from operations is not sufficient to meet required outflows.

In sum, financial distress is the events preceding and including bankruptcy, such as violation of loan contracts.

55. Financial Distress Costs

Legal and administrative costs of liquidation or reorganization (direct costs); an impaired ability to do business and an incentive toward selfish strategies such as taking large risks, underinvesting, and milking the property (indirect costs).

56. Financial Engineering

Creating new financial instruments by combining other derivatives or more generally, by using derivatives pricing techniques.

57. Financial Futures Contract

A commitment between two parties to exchange a standardized financial asset through an organized exchange at a specified price of futures contracts changes prior to delivery, and participants must settle daily changes in contract value.

58. Financial Innovation

The continuous development of new products, services and technology to deliver products and services.

59. Financial Intermediaries

An area of finance that deals with financial institutions, such as banks and insurance companies, which collect funds from savers and lend them to or invest them in businesses or people that need cash. Institutions that provide the market function of matching borrowers and lenders or traders. Financial institutions may be categorized as depository, contractual savings, and investment-type. Alternatively, they can be classified into depository institutions, insurance companies, securities firms and investment banks, mutual funds and finance companies.

60. Financial Investment

The net change in financial assets held by a sector or unit in the economy over a specified time period.

61. Financial Lease

A long-term noncancelable capital lease, generally requiring the lessee to pay all maintenance fees. Tax law identifies lessor as the owner of the leased asset, so the lessor can deduct the depreciation over the life of the lease. [See **Capital lease**]

62. Financial Leverage

Just as operating leverage arises from fixed operating costs, financial leverage arises from fixed financing costs. Financial leverage magnifies any change in EBIT to produce a percentage change in earnings per share larger than the change in EBIT. Financial leverage defines extent to which a firm relies on debt. Financial leverage is measured by the ratio of long-term debt to long-term plus equity.

63. Financial Management Analysis

Financial management analysis is a field in finance that studies how an organization should manage its assets, liabilities, and equity to produce a good or service.

64. Financial Market Integration

In integrated financial markets, domestic investors can buy foreign assets and foreign investors can buy domestic assets. Among countries that are fully integrated into world financial markets, assets with identical risk should command the same expected return, regardless of location. Integration of emerging market economies

into the world financial markets is generally followed by: a significantly larger and more liquid equity market; stock returns which are more volatile and more correlated with world market returns; a lower cost of capital; improved credit ratings; real exchange rate appreciation; and increased economic growth. They find strong evidence of structural breaks in emerging equity markets, but no evidence of structural breaks in the world equity market. However, these breaks do not always correspond very closely to the dates of official capital market reforms. For example, developed country financial markets are more integrated with worldwide financial markets than that of less developed countries.

65. Financial Markets

Markets that deal with cash flows over time, where the savings of lenders are allocated to the financing needs of borrowers. Financial markets are composed of money markets and capital market. [See **Money** and **Capital markets**]

66. Financial Mathematics

Financial mathematics describes the application of mathematics and mathematical modeling to solve financial problems. It is sometimes referred to as quantitative finance, financial engineering, and computational finance. The discipline combines tools from statistics, probability, and stochastic processes and combines it with economic theory.

67. Financial Planning

Financial planning is the process of analyzing alternative investment, financing, and dividend strategies in the context of various potential economic environments. Planning involves forecasting both the outcomes of different strategies and their risks. Thus, financial planning models are tools to help managers improve their forecasts of important accounts of financial statements and better understand the interactions of investment, financing, and dividend decisions.

Planning involves using different economic and sales scenarios and reacting to them with different strategies. Playing what-if games helps managers select an optimal course of action, given managers' risk preferences and beliefs about the most likely scenarios.

In developing a long-term financial plan, these three decisions (policies) can be described more explicitly as follows:

1. *The firm's investment decision.* This refers to the amount of cash needed for the firm's investment in a new asset (it is also called the *capital budgeting decision*). In addition, it also refers to the amount of working capital needed on an ongoing basis (also referred to as the *working capital decision*).
2. *The firm's financing decision.* This refers to new borrowing or new equity issued for financing the firm's investment in new assets. This decision is influenced by

the degree of financial leverage the firm chooses to employ and how it plans to raise the necessary new funds.

3. *The firm's dividend decision.* This refers to the amount of cash the firm thinks is necessary and appropriate to pay equity holders as cash dividends.

At the most basic level, a planning model is a tool that uses inputs supplied by managers in the form of economic, accounting, market, and policy information.

68. Financial Ratios

Financial ratios are created with the use of numerical values taken from financial statements to gain meaningful information about a company. The numbers found on a company's financial statements – balance sheet, income statement, equity statement, and cash flow statement – are used to perform quantitative analysis and assess a company's liquidity, leverage, growth, margins, profitability, rates of return, valuation, and more. Financial ratios can be classified into five groups. They are liquidity ratios, capital structure ratios, turnover ratios, profitability ratios, and market value ratios.

69. Financial Reform

Wall Street Reform or Financial Reform refers to reform of the financial industry and the regulation of the financial industry in the United States. Wall Street is the home of the country's two largest stock exchanges, and "Wall Street" is a metonym for the American financial sector. Major Wall Street reform bills include the Federal Reserve Act of 1913, the Glass-Steagall Act of 1933, the Truth in Lending Act of 1968, the Community Reinvestment Act of 1977, the Gramm–Leach–Bliley Act of 1999, and the Sarbanes-Oxley Act of 2002. The most recent Wall Street reform bill, the Dodd–Frank Wall Street Reform and Consumer Protection Act, was signed by the President of the United States Barack Obama on July 22, 2010, following a global financial crisis.

70. Financial Requirements

In the financial plan, financing arrangements that are necessary to meet the overall corporate objective. The plan will include a section on financing arrangements. This part of the plan should discuss dividend policy and debt policy. Sometimes firms will expect to raise equity by selling new shares of stock. In this case the plan must consider what kinds of securities must be sold and what methods of issuance are most appropriate.

71. Financial Risk

Financial risk measures the additional risk that the firm's stockholders bear when the firm is financed with debt as well as equity.

Financial risk is determined by how the firm decides to finance its assets. Financial risk occurs as a result of fixed costs in a firm's financial structure. A firm's financial structure is the combination of debt and equity that it uses to finance assets. Equity dividends, including preferred stock dividends, are considered to be a variable financing cost, as the firm can reduce the dollar amount of dividends or eliminate them entirely if its cash flow is poor. Shareholders may be unhappy, but even preferred shareholders can do little to force the firm to pay dividends. In sum, financial risk refers to potential variation in income before interest and taxes associated with fixed interest payments on debt and lease payments.

72. Financial Services Holding Company

A parent company that owns a bank holding company plus other subsidiaries, such as a thrift holding company and insurance subsidiary.

73. Financial Statement Analysis

Financial statement analysis is the process of analyzing a company's financial statements for decision-making purposes. External stakeholders use it to understand the overall health of an organization as well as to evaluate financial performance and business value. Financial ratio analysis is the main tool of financial statement analysis, using the five types of financial ratios to evaluate the company.

74. Financial Statistics

Financial statistics include all numerical data that summarizes past behavior or forecasts future behavior of an individual financial security, a group of securities, or markets in a broad geographic region. It is useful to first categorize these statistics into one of three areas. Global or national statistics examine the behavior of business or economic conditions that influence the value of securities in general. Market or industrial statistics track the activity of a specific set of securities linked by a common trading market or industry classification. Company-specific statistics examine performance of individual firms.

75. Financial System

The collection of markets, individuals, institutions, laws, regulations, and techniques through which bonds, stocks, and other securities are traded, financial services produced and delivered, and interest rates determined.

76. Financial Technology

Financial technology (Fintech) is used to describe new tech that seeks to improve and automate the delivery and use of financial services. At its core, Fintech is utilized to help companies, business owners, and consumers better

manage their financial operations, processes, and lives by utilizing specialized software and algorithms that are used on computers and, increasingly, smartphones. It is well known that machine learning and deep learning are important computer tools for financial technology.

77. Financial Z Score

[See **Credit scoring model**]

78. Financing Costs

Financing costs are defined as the interest and other costs incurred by the company while borrowing funds. They are also known as “Finance Costs” or “borrowing costs.” A company funds its operations using two different sources, Equity Financing and Debt Financing.

79. Finite Difference Method

A method for solving a differential equation. It can be classified into implicit, explicit, or other finite difference method.

80. Firm

A firm is a collection of assets, and the value of those assets depends upon the size, timing, and risk of their cash flows. Of all the possible goals of a firm, only shareholder wealth maximization fully considers the size, timing, and risk of the cash flow generated by the firm’s activities.

A firm purchases its assets with funds obtained from sources listed on the right-hand side of the balance sheet – liabilities and owners’ equity. Thus, the value of a firm belongs to its creditors and owners. Creditors have a fixed claim on the firm that does not change with variations in the value of the firm’s assets over time. As the shareholders have a residual claim on the firm’s assets, variations in a firm’s value are reflected mainly in the fluctuating value of the owners’ or shareholders’ wealth in a firm. Alternatives that increase shareholders’ wealth should be chose; alternatives that harm shareholders’ wealth should be rejected.

Fluctuations in the value of a firm are most easily seen in fluctuations in the market value of the shareholders’ claim on the firm. All else being constant, increases in shareholder value lead to a larger cushion for those with fixed claims on the firm, such as creditors, bondholders, employees, and pensioners.

Economics teaches that the goal of a firm is to maximize its economic profit, which is a function of the difference between the return earned on its assets and the opportunity cost of buying those assets. The workings of the financial markets will ensure that the cost to a firm of raising capital is equal to the capital’s opportunity cost; otherwise, available funds will flow to other firms that can offer investors higher expected returns at lower risk. Should returns earned by the firm exceed this

cost, the profit belongs to the firm's owners. Thus, the financial goal of maximizing shareholder wealth is similar to the concept of maximizing economic profit.

81. Firm Commitment Offerings

Investment banks distribute most IPOs in firm commitment offerings. With a firm commitment offering, the investment bank commits its capital to purchase IPO shares. Once the offering price is set, the bank purchases the shares at the offer price less a spread, or discount. The bank then sells the securities to investors. In practice, the investment bank lines up a number of investors to purchase the shares before the offering date. The spread represents the investment bank's profit from reselling each share at the offering price.

The issuer has virtually zero price risk in a firm commitment offering once the offer price is set. The issuer receives the proceeds from the sale immediately, which it can then spend as outlined in the prospectus. The investment bank carries, or underwrites, the risk of fluctuating stock prices. Should the market's perception of the issuer change or a macroeconomic event result in a stock market decline, the investment bank carries the risk of loss, or at least the possibility of a smaller than expected spread.

For most firm commitment underwritings, the managing investment bank arranges investment banking **syndicates** to help distribute shares of the newly public firm. The managing investment bank makes a smaller spread, or profit, from selling shares to syndicate members. [See also **Syndicates**]

82. Firm Commitment Underwriting

An underwriting in which an investment banking firm commits to buy the entire issue and assumes all financial responsibility for any unsold shares. [See **Firm commitment offerings**]

83. Firm-Specific Risk

[See **Diversifiable risk**]

84. First Mortgage Bond

A first mortgage bond has a primary, or senior, claim on assets. In theory, a first mortgage claim means that the underlying asset can be sold and the proceeds distributed to the first mortgage bondholders to satisfy their claims against the firm; any remaining funds from the sale are distributed to satisfy the **second mortgage** holders' claims. [See also **Second mortgage bond**]

85. First-Pass Regression

A time series regression to estimate the betas of securities or portfolios.

86. Fiscal Agent

A role of the Federal Reserve System in which it provides services to the federal government, such as clearing and collecting checks on behalf of the US Treasury and conducting auctions for sale of new Treasury securities.

87. Fiscal Policy

Fiscal policy involves planning government spending and taxing to influence economic conditions. Both tax laws and government expenditures affect the disposable income of consumers and corporations and, therefore, the level of aggregate demand in the economy. For example, taxes affect the incentives that people have to save and invest, and thus affect future economic growth. Tax laws affect firms' after-tax returns on their investments and thus help determine how a firm will invest today in order to generate future cash flows.

88. Fisher Effect

[See **Nominal risk-free interest rate**]

89. Fisherian Relation

The nominal interest rate in every contract will be equal to the real rate of interest plus the expected future inflation rate is called Fisherian relation:

$$(1 + R_j^t) = (1 + r_j^t)(1 + I_j^t)$$

where:

r_j^t = the real rate of interest in country j at time t

R_j^t = the nominal rate of interest at time t

I_j^t = the inflation rate at time

The implication of this relationship is that if the real rate of interest is equal everywhere, then the inflation differential between countries is fully reflected in their nominal interest rate.

90. Fixed Annuities

Annuity contracts in which the insurance company pays a fixed dollar amount of money per period.

91. Fixed Asset

Long-lived property owned by a firm that is used by a firm in the production of its income. Tangible fixed assets include real estate, plant, and equipment. Intangible fixed assets include patents, trademarks, and consumer recognition.

92. Fixed Asset Turnover Ratio

[See **Asset management ratios**]

93. Fixed-Charge Coverage Ratio

Ratio of earnings to all fixed cash obligations, including lease payments and sinking fund payments. [See **Capital structure ratios**]

94. Fixed-Income Security

A security such as a bond that pays a specified cash flow over a specified period.

95. Fixed Costs

Fixed cost is a cost that is fixed in total for a given period of time and for given volume levels. It is not dependent on the amount of goods or services produced during the period.

A factor affecting business risk is the firm's fixed costs. Fixed costs, such as rent, lease payments, and depreciation, remain the same whether the firm's sales, production, or profitability levels rise or fall. The effect of fixed costs on the firm's operating structure is to magnify, or leverage, the impact of a change in sales on EBIT. [See **Business risk**]

96. Fixed-Dollar Obligations

Conventional bonds for which the coupon rate is set as a fixed percentage of the par value.

97. Fixed Rate

An interest rate that does not change during a specified period of time. Fixed rate mortgage is a good example for this case.

98. Flat Benefit Formula

Method used to determine a participant's benefits in a defined benefit pension plan by multiplying months of service by a flat monthly benefit. [See **Defined benefit plans**]

99. Flat Volatility

The name given to volatility used to price an interest rate cap when the same volatility is used for each caplet. If different volatility is used for each caplet, then it is called spot volatility.

100. Flex Option

An option traded on an exchange with terms that are different from the standard options trades by the exchange.

101. Flight to Quality

Describes the tendency of investors to require larger default premiums on investments under uncertain economic conditions.

102. Float

Bankers define float as cash obligations that are in the process of collection. Another way to think of float is the difference between the balance shown in a firm's (or an individual's) checkbook and the balance on the bank's books. For instance, suppose that, on average, a firm writes \$10,000 worth of checks each day. If it takes 5 days for these checks to clear and be deducted from the firm's bank account, then the firm's own checking records will show a daily balance of \$50,000 lower than the bank's records. Conversely, if the firm, on average, receives \$10,000 worth of checks each day but deposits and clears these checks in only 3 days, the firm's books will show a balance \$30,000 higher than the balance on the bank's records. The difference between the \$50,000 negative float and the \$30,000 positive float, $-\$20,000$, is called the firm's *net float*. This suggests the possibility that a firm could consistently maintain a negative cash balance on its books, as long as it could accurately forecast its positive and negative clearings.

Float management is an integral component of the cash management system. To understand how to analyze and forecast float, we need to look at the five different types of float:

1. **Invoicing float** is the time it takes for a firm to bill receivables. The efficiency of the company's internal accounting and billing procedures affect this type of float.
2. **Mail float** is the time the firm's bill spends in the mail on its way to the customer and the time the customer's check spends in the mail on its way to the firm.
3. **Processing float** is the time between a firm's receipt of a payment and its deposit of the check for collection.
4. **Collection float** is the time from when the bank accepts a check for deposit to when it makes the funds available in the firm's checking account.
5. **Disbursing float** is the time between when a firm writes a check on available bank account funds and when the bank deducts the corresponding dollar amount from the firm's bank balance.

The first four components of float hinder the firm's ability to turn collection items into cash; these are examples of negative float. The fifth component, disbursing float, is positive float because it increases the amount of cash the firm has to use. High interest rates increase the benefits of reducing negative float or increasing net float.

Mail float generally is hard to control, but it can be controlled to some degree through the use of different collection sites. Processing and invoicing float result from internal company operations, so they can certainly be monitored and fine-tuned for increased efficiency. Collection and disbursement float can be reduced through cash collection and disbursement services provided primarily by the banking system.

In sum, float is the difference between bank cash and book cash. Float represents the net effect of checks in the process of collection, or clearing. *Positive float* means the firm's bank cash is greater than its book cash until the check's presentation. Checks written by the firm generate *disbursement float*, causing an immediate decrease in book cash but no change in bank cash. In *neutral float position*, bank cash equals book cash. Checks written by the firm represent *collection float*, which increases book cash immediately but does not immediately change bank cash. The sum of disbursement float and collection float is *net float*.

103. Floater

Floating-rate bond.

104. Floating-Rate Bond

A debt obligation with an adjustable coupon payment.

105. Floating Rate

An interest rate tied to a base rate that changes over time as market conditions dictate.

106. Floating-Rate Note (FRN)

A short-term note whose interest payment varies with a short-term interest rate.

107. Floating Lien

A floating lien gives the lender a claim against all the borrower's inventory without listing or specifying individual items. Such an arrangement makes it difficult, however, for the lender to prevent the borrower from running down inventories to a level that gives no real security for the loan; finance companies therefore are usually willing to advance only a small fraction of the estimated market value of the inventory against a floating lien.

108. Floor

An option position that guarantees a minimum price.

109. Floor-Ceiling Agreement

[See **Collar**]

110. Floor Broker

A licensed member of the exchange who is paid a fee for executing orders for clearing members of their customers.

111. Floor Plan Loans

Floor plan loans finance equipment purchases in an arrangement similar to a revolving credit agreement. Many manufacturers or distributors of machine tools, tractors, and similar heavy equipment supply these items to retailers under a floor plan system, which allows the retailer to pay for the merchandise only after actually selling it. The retailer's inventory therefore is financed by the supplier, either a manufacturer or a distributor. The manufacturer or distributor in turn finances this inventory by setting up a credit arrangement with a bank. Under such an arrangement, the bank pays the manufacturer for the equipment as soon as it is shipped. The bank then becomes the official owner of the equipment. When the equipment is sold, the retailer pays the wholesale price plus an interest charge directly to the bank. Alternatively, the retailer may give the manufacturer or distributor a note for the wholesale price of the equipment, which the manufacturer or distributor may then sell to the bank at a discount. This agreement compensates the bank, not by interest payments, but by the difference between the discounted sum it pays to the manufacturer and the full wholesale price it eventually will recover from the retailer.

112. Floorlet

One component of a floor.

113. Floor Rate

The rate in an interest rate floor agreement.

114. Floor Trader

An exchange member of the exchange who is paid a fee for executing orders for clearing members or their customers.

115. Flotation Costs

The firm cannot costlessly arrange to borrow money, either from a bank or by selling bonds or shares of stock. It costs money to raise money. The costs of issuing securities, flotation costs, include bank application fees; "points" paid on loans; the

accounting, legal, and printing costs of offering securities to the public; and any commissions earned by the investment bankers who market the new securities to investors. As a result of these costs, if the firm raises \$100 of funds, it actually receives less than \$100 to apply to the capital budgeting project. Thus, it must evaluate the cost of financing the project, net of issuing or flotation costs.

116. Flow of Funds Accounts

A system of social accounts prepared quarterly by the Board of Governors of the Federal Reserve System that reports the amount of savings and borrowing in the US economy by major sectors and how those savings and borrowings are allocated to different kinds of financial instruments.

117. Flow of Funds Matrix

The combined sources and uses of funds statements for all sectors in the economy as reported in the Federal Reserve Board's Flow of Funds Accounts.

118. Flower Bond

Special Treasury bond (no longer issued) that may be used to settle federal estate taxes at par value under certain conditions.

119. FNMA

Federal National Mortgage Association (Fannie Mae); a private corporation operating with an implicit federal guarantee; buys mortgages financed by mortgage-backed securities. [See **Fannie Mae**]

120. Forced Conversion

If the conversion value of a convertible is greater than the call price, the call can be used to force conversion.

121. Foreclosure

Selling property in order to apply the proceeds in payment of a debt.

122. Foreign Bank Debt

Foreign debt is money borrowed by a government, corporation or private household from another country's government or private lenders. Foreign debt also includes obligations to international organizations such as the World Bank, Asian Development Bank (ADB), and the International Monetary Fund (IMF). Total foreign debt can be a combination of short-term and long-term liabilities. Foreign

banking organizations (FBOs) operate a variety of banking institutions in the United States: branches, agencies, and US- or state-chartered Edge Act and Agreement corporations and banks. Foreign banking institutions play an important role in the US financial system; in December 2006, foreign banking institutions held over \$1 trillion in assets, approximately 11% of the total commercial banking assets in the United States.

123. Foreign Bank Relationships

Because foreign banking institutions play an integral role in the US financial system, they are supervised and regulated by US banking authorities. The types of regulations faced by a particular foreign banking institution depend in part on whether its US units are chartered in the United States or abroad. Foreign bank branches and agencies are legal extensions of their parent companies, and not freestanding entities in the United States. They do not have any capital of their own and face somewhat different regulations from other depository institutions in the United States. Edge and Agreement corporations and foreign banks' US subsidiaries, on the other hand, are freestanding legal entities with US or state charters and their own capital. Before the International Banking Act of 1978 (IBA) was passed, rules governing foreign banking institutions were significantly different from those governing US banking institutions. The IBA brought the different rules into closer alignment, particularly those relating to chartering, branching, and reserve requirements. The Depository Institutions Deregulation and Monetary Control Act (DIDMCA) of 1980, continuing in the same direction as the IBA, included provisions granting foreign bank branches and agencies operating in the United States direct access to Federal Reserve services and privileges such as check clearing, provision of coin and currency, Fedwire, and the discount window. Federal Reserve services and privileges are offered to foreign bank branches and agencies under the same rules and at the same prices as other depository institutions. Another important provision of DIDMCA was to subject all foreign banking institutions accepting deposits to Federal Reserve requirements.

124. Foreign Bonds

An international bond issued by foreign borrowers in another nation's capital market and traditionally denominated in that nation's currency.

125. Foreign Currency Option

An option on a foreign exchange rate. The valuation model for the European type of currency call option can be defined as:

$$C = Se^{-rfT}N(d_1) - Xe^{-rT}N(d_2)$$

where:

S = spot exchange rate

r = domestic risk-free rate

r_f = foreign risk free rate

X = exercise price

σ = standard deviation of spot exchange rate

$$d_1 = \left[\ln\left(\frac{S}{X}\right) + \left(r - r_f + \frac{\sigma^2}{2}\right)T \right] / \sigma\sqrt{T}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

126. Foreign-Currency Futures

A foreign-currency futures contract is similar to other commodity-futures contracts. It promises future delivery of a standard amount of a foreign currency at a specified time, place, and price.

127. Foreign Exchange

Currency of a foreign country acceptable as a medium exchange.

128. Foreign Exchange Brokers

Business firms that, in return for a commission, bring buyers and sellers of various currencies together to carry out currency trading.

129. Foreign Exchange Market

The foreign exchange market is not a geographic place; it consists of a communications network through which many participants throughout the world agree to buy or sell currencies. The foreign exchange market includes a wide variety of smaller markets for immediate exchanges (spot trading), agreements for later exchanges (forward trading), and contracts based on exchange rates (futures and options trading).

Given the worldwide dispersion of the foreign exchange market, exchange trading never opens or closes. Markets around the world are interconnected by communication links so that it is possible to trade in the foreign exchange market somewhere in the world 24 h a day, 7 days a week. This interconnection of diverse market segments also provides very competitive prices, which usually are within one-hundredth of a cent of each other.

130. Foreign Exchange Rates

The prices of foreign currencies expressed in terms of other currencies.

131. Foreign Exchange Risk

The risk that the value of a position dominated in a foreign currency may decline due to a change in exchange rate. For a financial institution (FI), it refers the risk that foreign exchange rate changes can affect the value of an FI's assets and liabilities located abroad.

132. Foreign Exchange Swap

An agreement to exchange stipulated amounts of one currency for another at one or more future dates.

133. Foreign Tax Credit

Income taxes paid to a foreign country that can be claimed as a tax credit against a domestic tax liability.

134. FOREX

Foreign exchange services offered by multinational banks that trade in various currencies and hold inventories of currencies for the convenience of their customers.

135. Forward Calendar

The anticipated supply of new bonds or other securities expected to come to market over the new week, month, or other calendar period.

136. Forward Contract

A forward contract is an agreement between a commercial bank and a corporate customer to exchange a specific amount of one currency for another on a specific future date at a specific price or exchange rate. At the initiation of the agreement, no money changes hands; the actual exchange of funds takes place on the future date specified in the forward contract. Forward contracts are very useful because they can be tailored to fit any situation, but they are very expensive.

137. Forward Curve

The set of forward or futures prices with different expiration dates on a given date for a given asset.

138. Forward Market

Channel through which currencies, securities, commodities, or other goods and services are traded for future delivery to the buyer with the terms of trade set in advance of delivery.

139. Forward Exchange Rate

The forward price of one unit of a foreign currency. Forward exchange rates can be used to control the risk of fluctuating spot rates over a specified time period. Forward rates allow a participant to “lock in” an exchange rate today for a transaction that will occur sometime in the future. In other words, forward exchange rate is a future day’s exchange rate between two major currencies. [See **Spot exchange rate**]

140. Forward Interest Rate

Rate of interest for a future period that would equate the total return of a long-term bond with that of a strategy of rolling over shorter-term bonds. The forward rate is inferred from the term structure.

141. Forward Parity

The relationship that forward exchange rate (F_{ij}^t) must be equal to the spot exchange rate at some point in time (S_{ij}^{t+1}) :

$$S_{ij}^{t+1} = F_{ij}^t$$

Forward parity must be true given the three relationships (interest-rate parity, purchasing-power parity, and Fisherian relation).

142. Forward Premium

The annualized percentage difference between the forward price and the spot price.

143. Forward Price

The delivery price in a forward contract that causes the contract to be worth zero.

144. Forward Rate

A forward rate is a rate quoted today on a forward loan that originates at some future period.

145. Forward Rate or Agreement (FRA)

Agreement that a certain interest rate will apply to a certain principal amount for a certain time period of the future.

146. Forward Risk-Neutral World

A world is forward risk-neutral with respect to a certain asset when the market price of risk equals the volatility of that asset.

147. Forward Start Option

An option designed so that it will be at the money at some time in the future.

148. Forward Strip

Another name for the forward curve.

149. Forward Swap

Used when new debt is to be issued at a future date; allows issuer to hedge against an undesirable increase in rates before the securities are issued. [See **Deferred swap**]

150. Forward Trade

An agreement to buy or sell based on exchange rates established today for settlement in the future. [See **Spot trade**]

151. Fourth Market

Direct trading in exchange-listed securities between one investor and another without the benefit of a broker.

152. Franchising

A method of producing and delivering financial services in which a financial institution sells the services it produces through locally licensed suppliers.

153. Francis and Rowell Model

The Francis and Rowell model is a financial analysis and planning model. This model is composed of 10 sectors: (1) industry sales, (2) production sector, (3) fixed capital-stock requirements, (4) pricing, (5) production costs, (6) income, (7) new financing required, (8) risk, (9) costs of financing, and (10) common stock valuation.

154. Free Cash Flow

Cash flow available after payment of all taxes and after all positive NPV projects have been provided for.

155. Frequency Distribution

The organization of data to show how often certain values or ranges of values occur. For example, a frequency distribution for either binomial or normal distribution.

156. FRMs

Home mortgage loans carrying an interest rate that is fixed for the life of the loan.

157. Full Capacity Unit Output

Full capacity unit output means a company's output reaches full capacity utilization.

158. Full-Service Broker

A brokerage that provides a full range of services to customers including advice on which securities to buy and/or sell.

159. Fully Diluted Earnings Per Share

Earnings per share expressed as if all outstanding convertible securities and warrants have been exercised. [See **Dilution**]

160. Fund Performance

There are many other factors to consider when analyzing a mutual fund's portfolio. By analyzing the sector weights of a fund and the fund manager's attributions to performance, an investor can better understand the historical performance of the fund and how it should be used within a diversified portfolio of other funds. An investor can also break down the portfolio into market cap groupings and determine whether the fund manager is particularly skilled at picking companies with certain size characteristics. The performance of mutual funds can be evaluated by the Sharpe performance measure, Treynor performance measure, and Jensen performance measure.

161. Fundamental Analysis

Research to predict stock value that focuses on such determinants as earnings and dividends prospects, expectations for future interest rates, and risk evaluation of the firm.

162. Fundamental Analysis

Fundamental analysis (FA) is a method of measuring a security's intrinsic value by examining related economic and financial factors. Fundamental analysts' study anything that can affect the security's value, from macroeconomic factors such as the state of the economy and industry conditions to microeconomic factors like the effectiveness of the company's management.

163. Fundamental Betas

Fundamental betas are estimates of future betas, based upon both industry- and firm-specific balance sheet and income statement data. Researchers have found that the average betas of different industries vary as a result of differences in their business risk. In addition, researchers have discovered that financial statement relationships are useful in predicting a firm's future beta. Betas change over time as a firm's growth, dividend-payout ratio, earnings variability, financial leverage, and size change. Studies have found that increased financial leverage and increased variability in sales and EBIT lead to larger betas, while higher dividend-payout ratios lead to lower betas.

164. Funding Decisions

In a corporation, sources of funds can be from equity or debt. Equity source of funds can be decomposed into internal source and external source. Internal sources of funds come from retained earnings and external funds come from equity issuance. Sources of debt can be borrowed from a bank or in issuing corporate debt. In addition, funding decisions can relate to dividend policy and financing policy. Dividend policy is related to the optimum payout ratio and financing policy is related to optimal capital structure.

165. Funding Requirements

Requirements that specify when funds need to be pumped into the project. They are derived from the cost baseline. Usually, funds are supplied in lump sum at specific times during the project. During the construction phase of a project, if part of the financing comes from debt, then it is necessary to simulate the impact of interests on the total funding requirement. This can trigger several modeling problems, the first of them being the creation of circular references if interests are calculated on the average of the opening and closing balances of debt.

166. Funds Flows

Funds flows reflect changes in various financial statement accounts and transfers of funds from one account to another.

167. Future Value

The future value represents the dollar amount that the current cash flow will come to be worth in the future if it earns interest (or grows) at a given rate over time.

Future Value After One Period

It is very straightforward to calculate the value of, for instance a \$100 investment after 1 year at a 10% annual rate of interest. The future value will (FV) be \$100 plus 10% of \$100, or \$110.

$$FV_1 = \$100 + (0.10)(\$100) = (\$100)(1 + 0.10) = \$110$$

Future Value After Two or More Periods

If the money is to be invested for 2 years, the value of the CD after that time will equal its value after 1 year plus an additional 10%:

$$FV_2 = \$110 + (0.10)(\$110) = (\$110)(1 + 0.10) = \$121$$

$$\text{or } FV_2 = (\$110)(1 + 0.10) = (\$100)(1 + 0.10)(1 + 0.10) = (\$100)(1 + 0.10)^2 = \$121$$

During the second year, the \$100 principal plus the first periods' interest of \$10 *both* earn interest. The \$10 of interest earned during the first year earns \$1 of interest (10% of \$10) during the second year. This growth illustrates the effect of compounding. For this reason, future value calculations are often called **compound value** calculations.

What if our CD can be rolled into a third year? That means the \$121 we have at the end of the second year will earn another 10%. At the end of the third year the CD will be worth:

$$FV_3 = (\$121)(1 + 0.10) = (\$100)(1 + 0.10)^2(1 + 0.10) = (\$100)(1 + 0.10)^3 = \$133.10$$

Over 3 years, the \$100 in principal has earned \$33.10 in interest.

These equations suggest a general formula for finding the future value (FV) in year n of a sum of money (PV):

$$FV_n = PV(1 + r)^n$$

The future value of \$PV is \$PV multiplied by a future value interest factor, $(1 + r)^n$. The future value interest factor, or FVIF, $(1 + r)^n$, will increase in size as either the interest rate or the number of years increases. It will increase exponentially as n increases, due to the effects of compounding interest. Higher interest rates will have a larger compounding effect.

Future Value of Several Cash Flows

The future values of several amounts are additive *if the amounts are paid at the same future point in time*. If a problem has several present cash flows or investments, we can easily find their total future value simply by adding the individual future values at the same future time.

168. Futures Contract

Unlike a **forward contract**, a futures contract is a standardized financial instrument with a stated amount and specific maturity that is traded on an organized exchange and is resalable up to the close of trading or settlement date. Futures contracts tend to be smaller than forward contracts and are not as flexible in meeting hedging needs. [See also **Forward contract**]

169. Futures Contract

A futures contract is a legal agreement to buy or sell a particular commodity asset, or security at a predetermined price at a specified time in the future. Futures contracts are standardized for quality and quantity to facilitate trading on a futures exchange. The buyer of a futures contract is taking on the obligation to buy and receive the underlying asset when the futures contract expires. The seller of the futures contract is taking on the obligation to provide and deliver the underlying asset at the expiration date.

170. Futures Exchange

A futures exchange is the arena for the actual daily trading of futures contracts. The exchange is a nonprofit organization whose members include those allowed to trade on its floor. Members include individual traders, brokerage firms, and other types of institutions.

171. Futures Market

The underlying purpose for futures market is to allow investors to display their uncertainties about the future. Futures markets allow for the transfer of risk from hedgers (risk-averse individuals) to speculators (risk seeking individuals), a key element necessary for the existence of futures market is the balance between the number of hedgers and speculators who are willing to transfer and accept risk.

172. Futures Options

An option on a futures contract.

173. Futures Overlay

Converting an investment in asset A into the economic equivalent of an investment in asset B by entering into a short futures position on asset A and a long futures position on asset B.

174. Futures Price

The delivery price currently applicable to a futures contract.

1.7 G

1. GAAP

Generally Accepted Accounting Principles representing the standard rules and procedures that accountant follow when reporting financial information in limited states.

2. Game Theory

Game theory is the study of mathematical models of strategic interaction among rational decision-makers. It has applications in all fields of social science, as well as in logic, systems science and computer science. Originally, it addressed zero-sum games, in which each participant's gains or losses are exactly balanced by those of the other participants. In the twenty-first century, game theory applies to a wide range of behavioral relations, and is now an umbrella term for the science of logical decision making in humans, animals, and computers. In addition, game theory has been frequently used to analyze macroeconomic behavior and real earnings management.

3. Gamma

The change in delta when the price of the underlying asset changes by one unit. Based upon the call option formula defined in option pricing model. [See **Option pricing equation**] The mathematic result can be defined as:

$$\frac{\partial^2 C}{\partial S^2} = \frac{1}{S\sigma\sqrt{T}} N'(d_1) > 0$$

4. Gamma-Neutral Portfolio

A portfolio with a gamma of zero.

5. GAP

Dollar value of rate-sensitive assets (RSAs) minus the dollar value of rate-sensitive liabilities (RSLs). Another way of comparing RSAs and RSLs is the GAP ratio, defined as: $\text{GAP Ratio} = \frac{\text{RSAs}}{\text{RSLs}}$

6. GAP Management

A technique for protecting a bank's or other financial institution's earnings from losses due to changes in interest rates by matching the volume of interest-sensitive assets held to the volume of interest-sensitive liabilities taken on.

7. Gap Option

An option where the option owner has the right to exercise the option at strike price K_1 if the stock price exceeds (or, depending on the option, is less than) the price K_2 . For an ordinary option, $K_1 = K_2$

8. GARCH Model

Generalized autoregressive conditional heteroskedasticity (GARCH) is a model for forecasting volatility where the variance rate follows a mean-reverting process.

9. Garnishment

A court directive authorizing a bank to withhold funds from a borrower.

10. Garn-St Germain Depository Institutions Act

A law passed by the US Congress in 1982 to further deregulate the depository institutions sector, especially the services offered by nonbank thrift institutions, and to give the federal deposit insurance agencies additional tolls to deal with failing institutions.

11. General Break-Even Analysis

A generalized formula for break-even quantity, Q^* can be defined as

$$Q^* = \frac{FC + OCF}{p - vc}$$

where:

FC = fixed cost

vc = variable cost per unit

p = price per unit

OCF = Operation cash flow

12. General Cash Offer

A public issue of a security that is sold to all interested investors, rather than only to existing share-holders.

13. General Credit Controls

Monetary policy tools that affect the entire banking and financial system, such as open market operations or changes in the Federal Reserve's discount rates.

14. General Obligation Bonds

Municipal bonds secured by general fund (i.e., the full faith), credit, and taxing power of the issuing state or local government. [See also **Revenue bond**]

15. General Partnership

Form of business organization in which all partners agree to provide some portion of the work and cash and to share profits and losses. Each partner is liable for the debts of the partnership. [See also **Limited partnership**]

16. Generalized Wiener Process

A stochastic process where the change in a variable in each short time period of length δt has normal distribution with mean and variance, both proportional to δt .

17. Generally Accepted Accounting Principles

In the United States, Generally Accepted Accounting Principles (GAAP) are used as guidelines for financial statement reporting. It is a common set of accounting concepts, standards, and procedures by which financial statements are prepared. Besides GAAP, thrifts and insurance companies are also subjected to statutory accounting. [See also **Statutory accounting**]

18. Gentry-De La Garza Model

A way of analyzing receivable balances is to use the Gentry-De La Garza model (see Financial Management 1985, pp. 23–38). This model describes three reasons why accounts receivable balances may increase: sales pattern effects, collection experience effects, and joint effect. Sales pattern effects are increases in receivables due solely to increases in sales.

The increase in accounts receivable is partly a function of increasing sales (the sales pattern effect), partly a function of the deterioration in collections (the collection experience effect), and some combination of both (the joint effect).

The joint effect is explained as the increase in receivables due to a simultaneous deterioration in collections and increase in sales. For instance, some customers may have purchased the product with the assumption that they could pay late; this would both increase sales and slow collection. The joint effect is calculated by taking the difference between the current and past collection experience effects and then multiplying this difference by the difference between the current and past sales levels.

Finally, the sales pattern effect is quantified by taking the difference between the current and past sales levels multiplied by the prior collection experience effect.

The sum of the sales pattern effect, joint effect, and collection experience effect equals the difference between the receivables balances.

The advantage of the Gentry-De La Garza model is that it separates the increase in receivables into three quantifiable components. A financial manager then can see clearly if an increase in receivables results from faulty credit controls or if the increase is simply consistent with rising sales levels. Deteriorating collections might indicate that credit is granted too freely or that the company is not persistent enough in collecting overdue accounts. Both problems fall under the control of the credit department and are the primary responsibility of the credit manager.

19. Geometric Mean (also Called Geometric Average)

If historical, or ex-post, data are known, an analyst can easily compute historical average return and risk measures. If X_t represent a data item for period t , the arithmetic average $\bar{X}$, over n periods is given by:

$$\bar{X} = \frac{\sum_{t=1}^n X_t}{n},$$

The sum of the values observed divided by the total number of observation – sometimes referred to as the mean.

Alternatively, we can use the same data to calculate the geometric average rate of return as

$$\bar{X}_g = \left[\prod_{i=1}^N (1 + X_i) \right]^{\frac{1}{N}} - 1$$

In general, the geometric average is less than the arithmetic average, and while both measures are intuitively plausible, we are left with the problem of determining which average is more appropriate to calculate. To illustrate, suppose you purchase a stock for \$10 per share and at the end of the first year the price is \$20 per share; you have experienced a 100% return. At the end of the second year, the price has returned to \$10 per share; you have experienced a loss of 50%. Using the arithmetic average of our yearly returns of +100% and –50% gives an average return of 25%, $(+100 - 50)/2$, which is ridiculous. We started with a stock value of \$10 per share, and the value of the stock at the end of the second year was \$10 per share, so we actually received a return of 0%. This is the amount calculated for the geometric average, $\sqrt[2]{(2)(1/2)} - 1$. In fact, viewing these averages as estimates of return that actually result from holding stock, Blume (1974, pp. 634–638) has shown that a *mixed average* of these two quantities is generally preferable to the individual use of either one. The weights in this compromise estimate depend on the length of time the stock

is to be held. Specifically, for a holding period of T years, Blume recommends estimating the annual holding period return by a mixed mean as follows:

$$\bar{X}_m = \frac{N-T}{N-1} \bar{X}_a + \frac{T-1}{N-1} \bar{X}_g$$

where N = numbers of periods of data used to calculate the average rates of return and T = number of periods the investment is to be held. Notice that if the stock is to be held for just 1 year ($T = 1$), then $\bar{X}_m$ is just the arithmetic average. However, for holding periods longer than 1 year, some weight is given also to the geometric average. Since it is a weighted average of these quantities, $\bar{X}_m$ must lie between $\bar{X}_a$ and $\bar{X}_g$.

20. Geometric Brownian Motion

A continuous stochastic process, $x(t)$, in which the increments are given as $dx(t)/x(t) = \alpha dt + \sigma dZ$, where dZ is the increment to a Brownian motion driving the process.

21. Gibson Relation

Gibson relation describes relationship between actual levels of prices and yields as follows:

1. When the price is relatively high, so are interest rates.
2. When prices are low, yields also tend to be low.

22. Gilts

British and Irish government securities.

23. Ginnie Mae

Name referring to the Government National Mortgage Association. [See **GNMA**]

24. Glass-Steagall Act

The 1933 act that separated lending activities from investment banking activities at commercial banks by prohibiting commercial banks from underwriting corporate securities. Since 1987, the Fed now allows bank holding companies to expand their activities in securities underwriting through the special investment bank subsidiaries of commercial bank.

25. Global Bonds

The international bond market is increasingly ignoring national boundaries. A growing number of debt issues are being sold globally. In 1989, the World Bank was

the first issuer of global bonds; in 1993, over \$15 billion of global bonds were issued. Global bonds usually are denominated in US dollars. As they are marketed globally, their offering sizes typically exceed \$1 billion. In addition to the World Bank, issuers include the governments of Finland and Italy and corporations such as Matsushita Electric Industrial CO., Citicorp, First Chicago Corp., and Korea Electric Power Co.

26. Global Financial Market Integration

The global integration of capital markets is at once a principal driver of globalization and a hallmark of the increasingly globalized economy. Capital markets are settings in which buyers and sellers of different kinds of capital – foreign currencies, corporate securities, government bonds, bank loans – meet to negotiate prices. Global capital markets are now open for business 24 hours a day and, thanks to information technologies, transactions can be carried out from anywhere in the world in a matter of seconds. International capital flows now routinely exceed international trade flows by a ratio of 10 to 1. Within global capital markets, portfolio investment and short-term investment now surpass foreign direct investment (FDI) and bank lending. The cross-border integration of increasingly volatile and dynamic capital markets creates obvious challenges for governance.

27. Global Minimum Variance Portfolio

The lowest-variance portfolio achievable, given a population of securities.

28. Globalization

Tendency toward a worldwide investment environment, and the integration of national capital markets.

29. GNMA

Government National Mortgage Association (Ginnie Mae) – a government entity that buys mortgages for low-income housing and guarantees mortgage-backed securities issued by private lenders. GNMA was created by congress in 1968.

GNMA is a government-owned agency with two major functions. The first is sponsoring mortgage-backed securities programs by FIs such as banks, thrifts, and mortgage bankers. The second is acting as a guarantor to investors in mortgage-backed securities regarding the timely pass-through of principal and interest payments on their sponsored bonds. In other words, GNMA provides **timing insurance**. In acting as a sponsor and payment-timing guarantor, GNMA supports only those pools of mortgage loans whose default or credit risk is insured by one of three government agencies: the Federal Housing Administration (FHA), the Veterans Administration (VA), and the Farmers Home Administration (FMHA). Mortgage loans insured by these agencies target groups that might otherwise be disadvantaged in the housing market, such as low-income families, young families, and veterans.

As such, the maximum mortgage under the FHA/VA/FMHA-GNMA securitization program is capped.

30. Going Private

A technique where a **leveraged buyout** (LBO) can be used to take a firm out of public ownership and into private ownership. [See also **Leveraged buyout**]

31. Going-Private Transactions

Public owned stock in a firm is replaced with complete equity ownership by a private group. The shares are delisted from stock exchanges and can no longer be purchased in the open market.

32. Going Public

Going public offers several advantages to a firm and its current, private shareholders. First, selling stock publicly allows the firm to tap another source of capital: the public equity markets. Managers may decide to take the firm public because they need more capital than a private placement can provide. Or it may be cheaper to raise public equity than to undergo another round of financing from venture capitalists. Studies show that a firm should be profitable and raise at least \$10 million for an IPO to be cost-efficient. In addition, once a firm goes public, it can raise money periodically from the public markets by selling additional shares of stock.

Second, a certain prestige and publicity surrounds a firm that goes public and lists its shares on a stock exchange for trading. Third, shareholders may enjoy attractive capital gains if management achieves sales and profit goals. Founding entrepreneurs often purchase shares for pennies when the firm begins operating, but after the IPO, their shares are worth much more. A good time to go public is when investors favor stocks in the firm's industry or when the stock market is in a strong rising trend. The IPO market may be momentarily hot in a certain industry.

A public company enjoys a fourth advantage through its shares' liquidity. Since investors can buy or sell shares easily from each other, investors or managers easily can sell all or part of their investments if they choose. Managers may receive pressure to go public from the firm's private equity holders – especially venture capitalists – who may have a strong desire to liquidate their holdings. Secondary market liquidity eases owners' worries about receiving fair market value for their shares, since an impersonal marketplace, rather than accountants and attorneys, determine the per-share value of the company. Public trading may take the shares even more valuable, by reducing their liquidity risk.

Yet, some firms find going public an undesirable option. First, offering stock to the public is an expensive process. The costs of preparing financial statements, hiring attorneys, and marketing the shares to investors can consume a significant portion of the funds raised.

Another drawback is loss of control over the firm. Unless the firm offers less than 50% of its equity to the public, investors who are unknown to current managers and owners will collectively own most of the firm's common stock. Those shareholders will elect a board of directors to ensure that decisions are made in the shareholders' best interests. In addition, shareholders will make other major decisions themselves by voting, as outlined in the corporate charter or as allowed by the board. Additionally, since the former private shareholders will lose control over who buys the publicly traded shares, they may find the firm the target of a hostile takeover sometime in the future. Of course, control can be diluted by selling shares privately as well.

A third potential disadvantage is that a public firm must lay out its finances for all to see. While this reporting requirement allows current and potential investors to examine the firm's strengths and weaknesses and gain insight into management's future plans, it also allows the firm's competitors, both foreign and domestic, to do the same thing. Rivals can factor in the firm's profit margins and product sales as they play their marketing and R&D strategies.

Public firms also must submit to regulation by the **Securities and Exchange Commission (SEC)** and the exchange on which their shares are traded. [See also **Securities and Exchange Commission**]

Finally, having shares listed and traded on an exchange does not always guarantee a dramatic increase in liquidity and share price. If the firm is still relatively small and the market sees no spectacular potential for future growth in sales and profits, investors may ignore the firm after the public offering. This could leave shareholders with shares of a public firm that nobody else really wants to own. The shares can become illiquid quickly if they are not traded frequently, and weak interest can leave them languishing at a low price. The firm may have been better off staying private.

33. Gold Exchange Standard

A system for making international payments in which each national currency is freely convertible into gold bullion at a fixed price and also freely convertible into other currencies at relatively stable prices.

34. Gold Standard

A monetary system where the value of a country's currency is determined by the value of the gold content in the currency.

35. Golden Parachute

Companies often provide their top executives with substantial severance benefits, or golden parachutes, in the event of hostile takeovers. These benefits may have some economic justification because top managers face a dilemma when a hostile threat emerges. If they resist a successful takeover attempt, even if they believe resistance in shareholders' best interests, then they are likely to face dismissal by the

new owners. Without the golden parachute, managers may have some incentive to acquiesce too easily to strong, hostile suitors.

36. Goodwill

An intangible asset representing the difference between the book value of an asset or a firm and the actual sales price.

37. Gordon Model

Many firms have sales and earnings that increase over time; their dividends may rise, as well. If we assume that a firm's dividends grow at an annual rate of g percent, next year's dividend, D_1 , will be $D_0(1 + g)$. Generalizing,

$$D_t = D_0(1 + g)^t$$

Substituting this into the equation for the present value of all future dividends, we can show that the price at any future time t can be defined as:

$$P_t = \frac{D_{t+1}}{r - g};$$

where:

P_t = firm's stock price at time t

$D_{t+1} = D_t(1 + g)$, next year's expected dividend (equals the current dividend increased by g percent)

g = the expected (constant) dividend growth rate

r = required rate of return

This result, known as the Gordon model, or the **constant dividend growth model**, provides a straightforward tool for common stock valuation. The main assumption of constant growth in dividends may not be realistic for a firm that is experiencing a period of high growth or negative growth (i.e., declining revenues). Neither will constant dividend growth be a workable assumption for a firm whose dividends rise and fall over the business cycle. The constant dividend growth model also assumes a dividend-paying stock; the model cannot give a value for a stock that does not pay dividends. In addition, in the denominator of the equation, the required rate of return, r , must exceed the estimated growth rate, g . Finally, the constant dividend growth model assumes estimates for r , the required rate of return, and g , the dividend growth rate.

The constant dividend growth model reveals that the following three factors affect stock prices, *ceteris paribus*: (1) the higher the dividend, the higher the stock price; (2) the higher the dividend growth rate, the higher the stock price; (3) the lower the required rate of return r , the higher the stock price.

38. Government Intervention in the Foreign Exchange Market

Buying or selling currencies, domestic or foreign, by a government agency or central bank in order to protect the value of the home currency in international markets.

39. Government National Mortgage Association (GMNA)

A federal government agency created in 1968 to assist the nation's home mortgage market through such activities as purchasing mortgages to finance low-income family housing projects and guaranteeing principal and interest payments on securities issued by a private mortgage lenders that are backed by pools of selected home mortgages.

40. Government-Sponsored Agencies

Institutions originally owned by the federal government but now privately owned with the authority to borrow from and lend money to private businesses and individuals or to issue loan guarantees.

41. Grace Period

The time period for a credit card statement representing the time from when the statement is generated to the last day full payment can be made and still avoids a finance charge.

42. Grandfather Clause

A legislative provision that exempts parties previously engaged in activities prohibited by new legislation.

43. Greek Letters for Option Pricing

Current stock price per share, exercise price, volatility, time to maturity, and the risk-free rate are the five important variables in determining option pricing. Greek letters are used to represent the partial derivative of option values with respect to these five variables.

44. Greeks

A term generally referring to delta, gamma, vega, theta, and rho, all of which measure the change in the price of a derivative when there is a change in an input to the pricing formula.

45. Greenmail

Payments to potential bidders to cease unfriendly takeover attempts. Managers may arrange targeted repurchase forestall a takeover attempt. In a targeted repurchase, a firm buys back its own stock from a potential bidder, usually at a substantial premium. These premiums can be thought of as payments to potential bidders to delay or stop unfriendly takeover attempts. Critics of such payments label them greenmail.

46. Green Shoe Provisions

Some IPOs contain Green Shoe provisions, named after one of the first firms to include the provision in its underwriting agreement. A Green Shoe provision gives the leading investment bank the right to increase the number of shares sold in the IPO, typically by 10–20% of the original offering. This helps the investment bank satisfy more investors if demand for an issue is particularly hot. This also gives investment banks another way to increase their profits, since they earn the spread on any extra shares they sell.

47. Gross Domestic Product (GDP)

The market value of goods and services produced over a period of time including the sum of consumer expenditures, investment expenditures, government expenditures, and net exports (exports minus imports).

48. Growing Perpetuity

A constant stream of cash flows without end that is expected to rise indefinitely. For example, cash flows to the landlord of an apartment building might be expected to rise a certain percentage each year.

49. Growth Funds

Growth funds are structured to include a well-diversified combination of common stock. Basically, three reasons may be cited. First, empirical studies of common stock have almost invariably shown their long-term total return to exceed those on bonds. Second, stock is generally conceded to be a better hedge against inflation risk than bonds. Third, many small investors may prefer to hold obligations of financial institutions as their major fixed-income securities because of their convenience and safety resulting from government insurance programs.

50. Growth Opportunity

Opportunity to invest in profitable projects.

51. Growth Rate

Growth rates refer to the percentage change of a specific variable within a specific time period. For investors, growth rates typically represent the compounded annualized rate of growth of a company's revenues, earnings, dividends or even macro concepts, such as gross domestic product (GDP) and retail sales. Expected forward-looking or trailing growth rates are two common kinds of growth rates used for analysis.

52. Guarantee

Make oneself liable for the debts of another.

53. Guaranteed Insurance Contract

A contract promising a stated nominal rate of interest over some specific time period, usually several years.

54. Guaranteed Investment Contract (GIC)

A financial contract in which the writer of a policy agrees to pay a fixed amount at maturity after receiving a fixed, single premium up front.

55. Guardian

An individual or a trust department appointed by a court to manage a minor's property or personal affairs.

1.8 H

1. Haircut

The collateral, over and above the market value of the security, required by the lender when a security is borrowed.

2. Hansen-Jagannathan Bound

Hansen–Jagannathan bound is a theorem in financial economics that says that the ratio of the standard deviation of a stochastic discount factor to its mean exceeds the Sharpe ratio attained by any portfolio. This result applies, among others, the Cauchy–Schwarz inequality. The Hansen-Jagannathan (H-J) bound is a type of mean-variance frontier. The main contribution is that it allows us to say something about moments of the stochastic discount factor, which is unobservable, in terms of moments of returns, which can be (in principle) observed.

3. Hazard Model

A statistical technique for determining “hazard functions,” or the probability that an individual will experience an event within a particular time-period, given that the individual was subject to the risk that the event might occur. Since the hazard or risk is likely to change over time, it is necessary to decide upon the functional form by which the hazard may depend upon time and the different explanatory variables, although there may in fact be little information upon which to base this decision. The hazard model is frequently used to predict the probability of default for a company.

4. Hazard Rate

Measures probability of default in a short period of time conditional on no earlier default.

5. HDD

Heating degree days. The maximum of zero and the amount by which the daily average temperature is less than 65° F. The average temperature is the average of the highest and lowest temperatures (midnight to midnight). The Chicago Mercantile Exchange began trading weather futures and European options on weather futures in September 1999. The contracts are on the cumulative HDD and cooling degree days (CDD) for a month observed at a weather station

6. Hedge

The forward markets allow users to hedge, or reduce the risk of, adverse currency fluctuations. Hedging is taking a position in two or more securities that are negatively correlated (taking opposite trading positions) to reduce risk.

7. Hedge Fund

Hedge funds are a type of investment pool that solicits funds from (wealthy) individuals and other investors (e.g., commercial banks) and invests these funds on their behalf. Hedge funds, however, are not technically mutual funds in that they are subjected to virtually no regulatory oversight (e.g., by the SEC under the Securities Act and Investment Advisors Act) and generally take significant risk.

8. Hedge Fund Performance

Understanding the performance and risk characteristics of hedge funds can often be quite a bit more complex than a mutual fund or standard portfolio of stocks and bonds. Many hedge funds seek absolute returns rather than trying to beat an index like the S&P 500, and so performance must be judged accordingly and depending on the particular strategy. Risk, likewise, must be measured in ways that are compatible

with the investment goals, and may include value-at-risk (VaR) as well as analysis of fat tails.

9. Hedger

A market participant who has or will have a position in the cash commodity and who attempts to eliminate or reduce risk exposure by taking an offsetting position in the futures or forward market.

10. Hedge Ratio (for an Option)

The number of stocks required to hedge against the price risk of holding an option. Also called the option's delta. [See also **Delta**]

11. Hedge Ratio (for Futures)

The ratio of the size of a position in a hedging instrument to the size of the position being hedged.

$$\text{Hedge ratio} = \frac{\text{cov}(\Delta S, \Delta F)}{\text{var}(\Delta F)},$$

where ΔS (Change of spot) = $S_t - S_{t-1}$, ΔF (Change of future) = $F_t - F_{t-1}$

12. Hedging

Investing in an asset to reduce the overall risk of a portfolio.

13. Hedging Demands

Demands for securities to hedge particular sources of consumption risk, beyond the usual mean-variance diversification motivation.

14. Heston Model

An option pricing model in which the instantaneous variance of the stock return follows a mean-reverting square root process.

15. High-Frequency Trading (HFT)

High-frequency trading, also known as HFT, is a method of trading that uses powerful computer programs to transact a large number of orders in fractions of a second. It uses complex algorithms to analyze multiple markets and execute orders based on market conditions. Typically, the traders with the fastest execution speeds are more profitable than traders with slower execution speeds.

16. High-Ranked Analysts

Research analysts may not be above average in pure accuracy, but the winners they pick tend to win big. High-ranked analysts are known to be good at identifying winners, with over 70% of buy ratings correct, buys made up about 85% of all accurate ratings. Hold ratings performed worst, with only 20% of them playing out as expected.

17. High-Yield Bonds

Junk bond. [See **Junk bonds**]

18. Highly Leveraged Transaction (HLT)

Transaction in which borrower's debt increases sharply after the asset exchange, such as an LBO.

19. Historical Cost

The value for certain balance sheet items reflecting the original cost or amortized cost.

20. Historical Simulation

A simulation based on historical data.

21. Historical Volatility

The standard deviation of the continuously compounded return on an asset, measured using historical prices.

22. Holder-of-Record Date

The date on which holders of record in a firm's stock ledger are designated as the recipients of either dividends or stock rights. Also called date of record.

23. Holding Company

Besides negotiating terms or making a tender offer, parties to a business combination may form a holding company. A holding company is a corporation that owns sufficient voting stock of another firm (or several firms) to have effective control. This form of organization is quite common in the financial services and banking industries. Typically, the combination forms a new corporation and the shareholders of the firms to be combined exchange their old shares for shares of the new holding company. This type of transaction is advantageous because it can provide effective

control with as little as 10% or 20% of the outstanding stock, so a smaller investment is required. The holding company differs from other business combinations in that it can take advantage of legal loopholes in state and federal laws, including tax laws.

24. Holding Period

Length of time that an individual holds a security.

25. Holding-Period Return

The annualized rate of return expected or realized from holding a security over a specific period of time.

26. Holding-Period Yield

Holding-period yield (HPY) is a measurement of investment performance related to HPR. The HPY is the ratio of the change in the market value of the investment plus cash distributions received during the period divided by the original value of the investment. This is represented by:

$$HPY_t = (r_t) = \frac{(P_t - P_{t-1}) + C_t}{P_{t-1}} = \frac{P_t + C_t}{P_{t-1}} - 1$$

From this expression it is easy to see that HPY is equal to $HPR - 1$. The HPY defined in the equation is a discrete type of HPY. It assumes that the cash flows and investments occur at specific points in time.

The frequency of compounding influences the HPR and HPY calculations in the following way:

$$HPY_t^c = \ln \left(\frac{P_t + C_t}{P_{t-1}} \right)$$

where HPY_t^c is the holding-period rate of return with continuous compounding, and $\ln$ is the natural logarithm.

More generally, the rate of return with continuous compounding for a given period is expressed by:

$$HPR_t^d = 1 + HPY^d = \exp(HPY^c)$$

where HPR_t^d is the discrete holding-period rate of return and $\exp(e)$ is 2.718, the base of natural logarithms. By taking the natural log of both sides of the equation:

$$\ln(1 + HPY^d) = HPY^c$$

In every case except $HPY^d = 0$, the continuously compounded return is always less than the discrete return

On the other hand, given a continuous return, the discrete return can be calculated using:

$$HPY^d = \exp(HPY^c) - 1$$

27. Holiday Calendar

Calendar defining when days are holidays for the purposes of determining payment dates in a swap.

28. Home Banking

Actions involving the conduct of banking business taking place in customer's homes, including telephone and computer transactions.

29. Home Currency Approach

The home currency approach is a method for evaluating overseas projects. This technique converts foreign currency cash flows to the home currency of the parent firm. Assuming that the home currency is the US dollar, it then discounts the US dollar cash flows at the project's US minimum required return to find the net present value.

The financial analyst can rely on forecasting services that analyze relative economic and political trends to predict future spot rates. Once the foreign cash flows are converted to dollars, the NPV calculation using the project's US required return is straightforward.

The steps in the home currency approach are summarized as:

1. Estimate foreign currency cash flows.
2. Predict future spot exchange rates using forecasts.
3. Convert foreign currency cash flows to home currency cash flows.
4. Compute project NPV using the project's required return.

30. Home Debit

A check drawn on a bank that is presented to the same bank for deposit or payment.

31. Home Equity Loan

Loan secured by an individual's equity in a home.

32. Homemade Dividends

An individual investor can undo corporate dividend policy by reinvesting excess dividends or selling off shares of stock to receive a desired cash flow.

33. Homemade Leverage

Idea that as long as individuals borrow (and lend) on the same terms as the firm. They can duplicate the effects of corporate leverage on their own. Thus, if levered firms are priced too high. Rational investors will simply borrow on personal accounts to buy shares in unlevered firms. [See **Modigliani and Miller (M&M) Proposition I**]

34. Homogeneous Expectations

The assumption that all investors use the same expected returns and covariance matrix or security returns as inputs in security analysis.

35. Horizon Analysis

Interest rate forecasting that uses a forecast yield curve to predict bond prices. Yield curve is a two dimension graph to present the relationship between yield to maturity and maturity. [See also **Yield to maturity**]

36. Horizontal Acquisition

Merger between two companies producing similar goods or service. For example, a steel company buys another steel company.

37. Horizontal Combination

If two firms had performed similar functions in the production or sale of goods and services, then the business combination is said to be *horizontal*. Before a horizontal combination, the firms were, or at least had the potential to be, competitors.

38. Horizontal Spread

[See **Calendar spread**]

39. Hot Money

Funds that move between institutions quickly in search of higher yields or greater safety. The hot money can cause a country's financial crises.

40. Howard-D' Antonio Strategy

Using a mean-variance framework, the Howard-D'Anton (1984, pp. 101–112) assumes that the “agent” is out to maximize the expected return for a given level of portfolio risk. A hedge ratio and measure of hedging effectiveness are derived in which the hedger's risk and return are both explicitly taken into account. The strategy can be expressed as:

$$\text{Hedge ratio } H = \frac{(\lambda - \rho)}{\gamma\pi(1 - \lambda\rho)}$$

$$\text{And hedging effectiveness } HE = \sqrt{\frac{1 - 2\lambda\rho + \lambda^2}{1 - \rho^2}}$$

where:

$\pi = \sigma_f/\sigma_s$ = relative variability of futures and spot returns

$\alpha = \bar{r}_f/(\bar{r}_s - i)$ = relative excess return on futures to that of spot

$\gamma = p_f/p_s$ = current price ration of futures to spot

$\lambda = \alpha/\pi = (\bar{r}_f/\sigma_f)/[\bar{r}_s - i)/\sigma_s]$ = risk-to-excess-return relative of futures versus the spot position

P_s, P_f = the current price per unit for the spot and futures, respectively

ρ = simple correlation coefficient between the spot and futures returns

σ_s = standard deviation of spot returns

σ_f = standard deviation of futures returns

$\bar{r}_s$ = mean return on the spot over some recent past interval

$\bar{r}_f$ = mean return on the futures over some recent past interval

i = risk-free rate

41. Hung Convertibles

Convertible bonds that have no chance of being converted are called hung convertibles. The idea here is that if the investors do not wish to convert their bonds into the firm's equity, the conversion process is hung up. The bond is worth more as a bond than it is worth converted into equity. APB No. 15 and FASB No. 55 require a firm to provide EPS information under either circumstance and let the market participant choose which measure is more meaningful.

42. Hybrid Security

A hybrid security is a security which has characteristics of both debt and equity. For example, convertible bond are securities that can be exchanged for a stipulated number of shares of common stock during a specific period.

43. Hypothecation

In a contract, committing property to secure a loan.

44. Hypothesis Testing

Hypotheses are assumptions about a population parameter. Hypothesis testing involves judging the correctness of the hypotheses. In fact, we often rely heavily on

sample data in decision making. For example, the results of public opinion polls may actually dictate whether a presidential candidate decides to keep running or to drop out of the primary race.

1.9 I

1. IBFs

International banking facilities located inside US borders that contain a set of computerized credit and deposit accounts offered by US multinational banks to their foreign customers.

2. Idiosyncratic Risk

An unsystematic risk. This risk can be diversified from a portfolio; hence it is also a diversifiable risk. [See **Unsystematic risk**]

3. Idiosyncratic Standard Deviation (Risk)

The idiosyncratic risk is the portion of risk that is unexplained by BETA. We calculate idiosyncratic volatility (IVOL) as the standard deviation of the residuals from a regression that uses Beta to estimate the relationship between a given asset and the market.

4. Illiquidity

Illiquidity is the opposite of liquidity; either an asset cannot be converted into cash (e.g., a leased machine cannot be sold to raise cash) or an asset cannot be sold at a reasonable price (e.g., a firm bought a machine for \$1 million, but the best offer from another buyer is \$100,000). In the latter case, if the firm keeps the asset and uses it, it is worth ten times more than the amount of cash it could raise in a sale in the market.

In the short run, many firms may be illiquid, that is, they may lack cash. They remedy this situation by short-term borrowing. A firm borrows cash to meet its current obligations, knowing that its cash flow will improve in the future. This kind of illiquidity is transitory and is not associated with insolvency or bankruptcy. On the other hand, if a firm faces illiquidity with no expectation of future cash flow improvement, illiquidity may lead to insolvency and bankruptcy.

5. Immunization

A strategy that matches durations of assets and liabilities so as to make net worth unaffected by interest rate movements. If interest rates rise, the present value of assets and liabilities will fall by the same amount. Similarly, if interest rates fall, then the value of the assets and liabilities will rise by the same amount.

6. Immunize

To fully hedge against interest rate risk. Alternatively, it refers that immunization occurs when a financial institution's equity holders are fully protected against interest rate risk.

7. Immunized

It describes a financial institution that is fully hedged or projected against adverse movements in interest rates (or other asset prices).

8. Impairment of Capital Rule

A legal constraint known as the impairment of capital rule is designed to protect the firm's creditors. It stipulates that dividends cannot exceed the amount of retained earnings listed on the balance sheet. This ensures that the firm retains enough capital to pay its legal obligations.

9. Implicit Agency Costs

[See **Agency costs**]

10. Implicit Contract

It is a long-term customer relationship between a borrower and lender based upon reputation. This kind of implicit contract is generally regarding borrowing and repayment that extends beyond the formal explicit legal contract.

11. Implied Distribution

A distribution for a future asset price implied from option prices.

12. Implied Forward Rate

The forward interest rate between t_1 and time t_2 ($t_1 < t_2$) that makes an investor indifferent between, on the one hand, buying a bond maturing at t_2 , and, on the other hand, buying a bond maturing at t_1 and reinvesting the proceeds at this forward interest rate.

13. Implied Rate Forecast

The market's expectation about future interest rates as indicated by the shape of the yield curve or by financial future prices.

14. Implied Repo Rate

The repo rate implied from the price of a Treasury bill and a Treasury bill futures price.

15. Implied Tree

A tree describing the movements of an asset price that is constructed to be consistent with observed option prices.

16. Implied Volatility Smile/Skew/Surface

Returns on financial securities (stocks, indexes, currencies) are not normally distributed. They all have fatter tails than normal (on average). The distribution is also skewed, mostly negative for stock indexes (and sometimes single name stocks) but can be either direction (positive or negative) for currencies. The return distribution is not constant over time but varies strongly. The volatility of the distribution is not constant and even higher moments (skewness, kurtosis) of the distribution are not constant, either. A good option pricing model should account for return non-normality and its stochastic (time-varying) feature.

17. Implied Volatility Spread

Implied volatility represents the current market price for volatility based on the market's expectations for future movement in a given underlying asset. This value is "implied" by the dollar and cent value of options trading in the marketplace. Aggregate implied volatility spread (IVS), defined as the cross-sectional average difference in the implied volatilities of at-the-money call and put equity options, is significantly and positively related to future stock market returns at daily, weekly, monthly, to semiannual horizons. This return predictive power is incremental to existing return predictors and is significant both in sample and out of sample. Furthermore, IVS can forecast macroeconomic news up to one year ahead. The return predictability concentrates around macro news announcement. Common informed trading in equity options offers an integrated explanation for the ability of IVS to predict both future stock market returns and real economic activity.

18. Implicit Finite Difference Method

A method for valuing a derivative by solving the underlying differential equation. The value of the derivative at time $(t + \delta t)$ is related to three values at time t . To pricing stock option, these three values are M = number of stock price; N = number of time maturity; and $(\Delta) S$ = stock of price intervals. For example, values 10, 5, and 3 were chosen for M , N , and $(\Delta) S$. Thus the option price is evaluated at \$3 stock price intervals between \$0 and \$30 and at half month interval between 0 and $2\frac{1}{2}$ months.

19. Implied Variance (Implied Volatility)

That state of the art method in the market today for estimating the volatility is the implied-variance estimate. Implied variance can generally be regarded as the market's opinion about the future variance of the stock. Originally proposed by Latane and Rendleman (1976, pp. 369–382), the idea behind the estimation of the implied variance is to equate the Black-Scholes model price to the current market price and solve iteratively for the remaining unknown variance. No closed-form solution is available to compute the implied variance, so a numerical search procedure such as the Newton–Raphson search or linear least-squares regression must be used. Issues abound concerning the use of implied variance; the first of them is the weighting issue. If the implied variance inherent in the market price for each outstanding option on a stock (underlying asset) were calculated, there would be as many estimates of the stock's implied variance as there are options. Disregarding the possibility of market mispricings for the moment, a number of other factors may also be able to help explain the observed discrepancies.

1. Exercise price (amount in or out of the money) differences
2. Time-to-maturity differences
3. Trading-volume differences among the individual options
4. Market conditions

20. Implied Volatility

The volatility for which the theoretical option price (typically computed using the Black-Scholes formula) equals the observed market price of the option. In other words, the expected volatility in return on an underlying asset or contract derived from an option pricing model.

21. Inception Profit

Profit created by selling a derivative for more than its theoretical value.

22. Income Beneficiary

One who receives income from a trust.

23. Income Bond

A bond on which the payment of income is contingent on sufficient earnings. Income bonds are commonly used during the reorganization of a failed or failing business.

24. Income Effect

The relationship between interest rate levels and the volume of saving in the economy which argues that the advent of higher interest rates may induce saver to save *less* because each dollar saved now earns a higher rate of return.

25. Income Fund

It provides liberal current income from investments. Income fund holds both equity and fixed-income security in a relatively stable proportion.

26. Income Statement

The income statement is an accounting report that summarizes the flow of a firm's revenues and expenses for a specific period. Unlike the balance sheet, it represents flow instead of static information. The income statement affects the balance sheet when the period's net income (or loss) less any dividends is added to (or subtracted from) retained earnings on the balance sheet. The income statement reports important information about the results of operations and gives reasons for the company's profits or losses.

The income statement may be produced annually, quarterly, or monthly. Company management uses monthly statements primarily for internal purposes, such as estimating sales and profit targets, controlling expenses, and monitoring the progress of long-term targets. Quarterly and annual income statements are especially useful to the firm's shareholders, creditors, and competitors. The top entry of the income statement gives net sales revenue. From this total, subsequent entries subtract expenses, such as the cost of goods sold, selling and administrative expenses, research expense, interest expense, and income tax expense. This gives the famous bottom line: net income.

Alternative accounting methods can also affect the size of reported net income. The methods for calculating depreciation, inventory value, and pension fund liabilities all influence the amount of reported profits. For example, historical cost accounting may understate the cost of goods sold; in an inflationary environment, this can result in an overstatement of sales, taxes, and net income.

Firms generally practice accrual accounting, recognizing revenues and matching corresponding expenses at the time of sale. Unless the firm sells its products only for cash, recognizing revenue does not mean that a cash inflow has occurred; cash will not flow into the firm until some time in the future, when the customer makes a payment on an account. Similarly, matching expenses to revenue distorts the perception of cash outflows. Firms must pay for many matched expenses, including raw materials production costs, and labor expenses, before they sell the corresponding goods. In addition, some income statement expense items do not reflect cash outflows, for example depreciation expense.

Thus, positive net income does not necessarily mean that cash inflows exceed cash outflows; neither does a negative net income figure imply imminent bankruptcy.

The analyst needs a better tool than an income statement to determine the cash flows of a firm, which is the purpose of the statement of cash flows. [See also **Statement of cash flows**]

In sum, income statement is a financial report that summarizes a firm's performance over a specified time period.

27. Income-and-Growth Funds

It is one kind of mutual funds. Income-and-growth funds are composed of a combination of common stock and bonds. Whether the emphasis is on income or growth determines what percentage of bonds or common stock is in the portfolio.

28. Incremental After-Tax Operating Cash Flows

For a cost-saving project, it is usually easier to estimate operating cash flows (OCF) by the tax shield approach:

$$\text{OCF} = (S - C)(1 - T) + T(\text{Dep}) - \text{Change in NWC}$$

where S = sales revenue, C = cost, T = tax rate, Dep = depreciation, NWC = net working capital.

The incremental sales revenue is expected to be zero for most cost-savings projects, including this one. Operating cash flows depend upon the estimated cost savings, the depreciation, the tax rate, and changes in net working capital.

29. Incremental Cash Flows

The stand-alone principle requires the analyst to examine the incremental cash flows that occur as a result of the project. The cash flows are incremental in that they represent the differences between the firm's after-tax cash flows *with* the project and its **base case**, or the after-tax cash flows *without* the project. [See also **Base case**]

Estimating incremental after-tax cash flows for a project requires a more thorough analysis than just determining the expected change in cash flows from the firm's current condition.

In sum, incremental cash flows are differences between the firm's cash flows with and without a project.

30. Indentures

All bonds will have indentures, which are contracts or agreements between issuing corporations and their bondholders. Such an agreement is supervised by a trustee who acts on the behalf of bondholders to ensure proper execution of the indenture provisions by the corporation. If the issuer violates indenture provisions, it is in default, and the trustee must act to protect the bondholders' interests.

31. Independent Bank

A bank operating in one locality that is not part of a large multibank holding company or group of banks.

32. Independent Projects

Projects are independent projects when acceptance or rejection of any one alternative would have no bearing on the acceptance or rejection of any other. The firm could undertake any or all of a group of independent projects, as long as each accepted project was expected to increase shareholder wealth.

33. Index Amortizing Swap

[See **Indexed principal swap**]

34. Index Arbitrage

An investment strategy that exploits divergences between actual future prices and their theoretically correct parity values to make a profit. [See **Program trading**]

35. Index Fund

A mutual fund holding shares in proportion to their representation in a market index such as the S&P 500.

36. Index Futures

A futures contract on a stock index or other index.

37. Index Model

A model of stock returns using a market index such as the S&P 500 to represent common or systematic risk factors.

38. Index of Leading Indicators

[See **Business cycle**]

39. Index Option

A call or put option based on a stock market index.

40. Index Rate

The rate that serves as a base rate when pricing certain mortgages and variable rate loans.

41. Indexed Principal Swap

A swap where the principal declines over time. The reduction in the principal on a payment date depends on the level of interest rates. (The lower the interest rate, the greater the reduction in the principal.)

42. Indifference Curve

A curve connecting all portfolios with the same utility according to their means and standard deviations.

43. Indirect Finance

Also known as financial intermediation, in which financial transactions (especially the borrowing and lending of money) are carried on through a financial intermediary such as a commercial bank.

44. Indirect Loan

Loan in which a retailer takes the credit application and negotiates terms with the actual borrower. The lender then purchases the loan from the retailer under pre-arranged terms.

45. Indirect Quotes

Exchange rate quotations can easily cause confusion. The market convention is to use indirect quotes; that is, a statement of units of foreign currency per US dollar (e.g., DM 1.4736/dollar). Thus, when economists expect the US dollar (\$) to strengthen against the yen (¥), they expect the indirect quote (¥/\$) of the exchange rate to rise, so the US dollar will purchase more Japanese yen. A weakening US dollar means the dollar will purchase less yen, so the indirect quote will fall.

46. Individual Retirement Account

A retirement account available to individuals to defer income taxes.

47. Industrial Development Bonds

Debt securities issued by a local government agency to aid a private company in the construction of a plant and/or the purchase of equipment or land.

48. Industrial Revenue Bond (IRB)

A bond issued by a state government, local government, or political subdivision for the express benefit of a business that will effectively use the proceeds.

49. Inefficient Market

[See **Efficient market**]

50. Inflation

When the economy begins to expand too quickly, demand from consumer and business spending may outstrip supply, driving prices upward; this is inflation. Although many argue that the Fed can best control inflation by slowing the growth of the money supply, others argue for lower levels of government spending and/or higher taxes to reduce aggregate spending in an overheated economy. In the case of economic recessions, some economists favor fiscal policies in the form of higher government spending or lower taxes in order to stimulate demand forces.

Differences in inflation rates between countries will lead to changes in the spot exchange rate over time. Countries with higher inflation rates will face depreciation, or increasing weakness, in their currencies over time.

51. Inflation Differential Risk

Inflation differential risk is the second added dimension of international diversification. Suppose an investor in the United States has a security in England whose return is fixed in terms of the pound. Assuming that there is no inflation in the United States but that the inflation rate in England is uncertain. The dollar value of the investment at the end of the period is uncertain and hence risky.

52. Inflation-Caused Depreciation Effect

Changes in the expected price inflation rate may not lead to equivalent increase in nominal interest rates due to the tendency of depreciation charges on existing plant and equipment to lag behind the rising cost of new replacement plant and equipment, discouraging business investment and credit demand.

53. Inflation-Caused Income Effect

The relationship between changes in the rate of price inflation and shifts in income (including consumption and saving) that lead to changes in real and nominal interest rates.

54. Inflation-Caused Income Tax Effect

The presence of a progressive income tax structure tends to cause nominal interest rates to increase by more than expected increase in the rate of inflation.

55. Inflation-Caused Wealth Effect

Changes in inflation expectations may alter the value of wealth held in financial assets by individuals and institutions, causing a change in their savings plans and leading to offsetting movements in real and nominal interest rates.

56. Inflation-Escalator Clause

A clause in a contract providing for increases or decreases in inflation based on fluctuations in the cost of living, production costs, and so forth.

57. Inflation Premium

The expected rate of price inflation which, when added to the real interest rate, equals the nominal interest rate on a loan.

58. Inflation Risk (or Purchasing Power Risk)

The probability that increases in the average level of prices for all goods and services sold in the economy will reduce the purchasing power of an investor's income from loans or securities.

59. Information Asymmetry

A type of transaction costs relates to the cost of information. Management's inside information is not freely disseminated. If management were to share its secrets and competitive plans with *The Wall Street Journal* or disclose them in mailings to shareholders, that information would quickly arrive in the offices of the firm's competitors.

Because of this information gap or information asymmetry between management and the public financial markets, the firm may need to raise money to take advantage of a competitive opportunity at a time when management feels that the firm's stock is underpriced. If raising equity is not an attractive option because of a low stock price and the firm's debt ratios already are high or are nearing the limits set by covenants in prior bond issues, the firm may miss the investment opportunity. Therefore, firms may avoid issuing excessive amounts of debt in order to maintain financing flexibility in the form of some degree of excess or unused debt capacity. In this way, the firm can maintain its ability to finance good capital budgeting projects by borrowing. This gives the firm another reason not to over-leverage itself; in fact, the firm may try to keep its debt ratio slightly below the optimal ratio in order to ensure that funds will be available if they are needed.

60. Information Revolution

A trend in the economy in which electronic innovation allows information to be moved faster and in greater volume between individuals and institutions.

61. Information-Content Effect

Payment of a dividend conveys a signal to the market place. In other words, each dividend payment or change in the dividend carries information to investors about the company. In sum, the rise or fall of stock price following the dividend signals is called information-content effect.

62. In-House Processing Float

Refers to the time it takes the receiver of a check to process the payment and deposit it in a bank for collection.

63. Initial Margin

The cash required from a futures trader at the time of the trade.

64. Initial Outlay

The first cash flow estimate is the initial investment in the project. For projects that require designing or modifying equipment and buildings, engineering estimates may be available. Engineers can examine preliminary designs or architectural sketches and estimate the quantities of various materials needed. Estimates of purchases, transportation costs, and construction expenses can be developed based on current market prices.

Another means of estimating the acquisition or construction cost of a project is to solicit bids from various construction or equipment manufacturers based upon a preliminary set of design specifications. An approximate cost can be determined through discussions with bidding firms. If the firm is large enough that it has an in-house engineering or real estate acquisition staff, this expertise also can be tapped to estimate relevant costs.

The expense of developing cost estimates is a sunk cost. That money is spent and gone whether or not the proposed project is accepted; it should not be included in the project's cash flow estimates. However, the initial outlay estimate must consider opportunity costs if the project will use property or equipment presently owned by the firm.

The investment cost estimate may have to be adjusted if the project involves replacing one asset with another, presumably newer and more cost-efficient model. If the old asset is going to be sold, the investment outlay must be reduced by the after-tax proceeds from the sale of the old asset.

Finally, even though a project's initial outlay may directly involve property and equipment (investing cash flows), it also may have implications for net working capital (operating cash flows). For example, if a project affects the firm's production process, inventory levels may change. New raw materials needs may affect accounts payable. These kinds of expected changes in net working capital must be included as part of the initial outlay.

65. Initial Public Offering

Raising capital privately may fail to raise the necessary funds or the cost of raising funds privately may be too high. When this happens, the firm may choose to go public in an initial public offering (IPO). That is, the firm may sell shares of stock to the general public and allow the shares to trade freely between investors. Many entrepreneurs dream fondly of their firms someday becoming public corporations in this way. [See also **Going public**] IPO is also called an unseasoned new issue.

66. Input List

List of parameters such as expected returns, variances, and covariances necessary to determine the optimal risky portfolio.

67. Inside Information

Nonpublic knowledge about a corporation possessed by corporate officers, major owners, or other individuals, with privileged access to information about a firm.

68. Insider Trading

Some managers may try to use their privileged access to private information about the firm for their own personal gain. By buying shares of stock before good news is announced and selling it prior to the release of bad news, insider trading allows them to profit inordinately as compared to the market as a whole. Such actions are illegal in the United States, but it is sometimes difficult to prove that an executive's stock purchases or sales were caused by his or her access to private information.

69. Insolvency

Insolvency means that the firm does not have sufficient cash inflows to meet all of its cash outflows.

Although all businesses expect to succeed, many do not. Various financial indications of serious difficulty often are apparent. Cash shortages may cause illiquidity, borrowing may increase, accounts may be overdrawn, and maintenance of plant and equipment may be delayed. Careful observation of either profit or cash receipt and disbursement trends may signal pending financial troubles. However,

frequently occurring illiquidity can make the difficulty so acute that the problem can no longer be ignored.

Cash flow problems can create either **technical** or **legal insolvency**. [See also **Technical insolvency** and **Legal insolvency**]

A firm that finds itself in financial distress due to a state of insolvency or failure to satisfy a bond indenture has several alternatives:

1. Do nothing, but hope something will come along to save the situation.
2. Attempt to sell out. The firm can try to find a buyer, but buyers of troubled firms may be few. Even if one can be found, the seller frequently feels fortunate to walk away with any portion of the original equity.
3. Seek adjustments with creditors outside the judicial process, commonly called a workout. Some arrangements between the firm and its creditors may permit it to keep operating with the hope that it can work its way out of trouble. Such adjustments usually take the form of extensions of repayment schedules and/or compositions of credit.
4. Seek court relief in bankruptcy proceeding in the form of a reorganization or liquidation.
5. Assign assets to a third party for liquidation.
6. Liquidate.

70. Insolvent

The financial position of a firm whose market value of stockholders' equity is less than or equal to zero. A firm is technically insolvent when the book value of stockholders' equity is less than or equal to zero.

71. Installment Credit

All liabilities of a borrowing customer other than home mortgages that are retired in two or more consecutive loan payments.

72. Installment Loan

A loan that is payable in periodic, partial installments.

73. Instantaneous Forward Rate

Forward rate for a very short period of time in the future.

74. Instruments

Financial securities, such as (1) money market instruments (e.g., commercial paper) or (2) capital market instruments (e.g., stocks and bonds).

75. Insurance Premium

An insurance premium is the amount of money an individual or business pays for an insurance policy. Insurance premiums are paid for policies that cover healthcare, auto, home, and life insurance. Once earned, the premium is income for the insurance company. It also represents a liability, as the insurer must provide coverage for claims being made against the policy. Failure to pay the premium on the part of the individual or the business may result in the cancellation of the policy.

76. Insurance Principle (the Law of Averages)

The average outcome for many independent trials of an experiment will approach the expected value of the experiment.

77. Interbank Loan

Credit extended from one bank to another.

78. Interest Coverage Ratio

Earnings before interest and taxes divided by interest expense. Used to measure a firm's ability to pay interest. [See **Capital structure ratios**]

79. Interest on Interest

Interest earned on reinvestment of each interest payment on money invested.

80. Interest Rate Cap

An option that provides a payoff when a specified interest rate is above a certain level. The interest rate is a floating rate that is reset periodically. The interest rate cap pays the difference between the realized interest rate in a period the interest cap rate.

81. Interest Rate Collar

A combination of an interest rate cap and an interest rate floor. The purchase of an interest rate collar is actually the simultaneous purchase of an interest rate cap and sale of an interest rate floor on the same index for the same maturity and notional principal amount.

82. Interest Rate Derivative

A derivative whose payoffs are dependent on future interest rate.

83. Interest Rate Floor

An option that provides a payoff when an interest rate is below a certain level. The interest rate is a floating rate that is reset periodically.

84. Interest Rate Insurance

An insurer agreeing to reimburse a borrower for additional interest expense if the borrower's loan rate climbs above some specified maximum loan rate.

85. Interest Rate Option

An option where the payoff is dependent on the level of interest rates.

86. Interest Rate Parity

Under interest rate parity, investors are indifferent between investing at home or abroad as far as expected return is concerned; any existing nominal risk-free interest rate disparity is offset by spot and forward exchange rate differentials. When interest rate parity exists, the following relationship is true.

$$S_0^*(1 + R_{FC})/F_1 = (1 + R_{US})$$

The left-hand side of the equation reflects the return from converting dollars at the spot rate (S_0), investing them at the foreign rate ($1 + R_{FC}$), and then converting the currency back into dollars at the forward rate (F_1). The right-hand side reflects the return from investing the dollars in the United States.

87. Interest Rate Risk

The general level of interest rates in an economy does not remain fixed; it fluctuates. For example, interest rates will change in response to changes in investors' expectations about future inflation rates. From the "**seesaw effect**," a rise in interest rates renders the fixed coupon interest payments on a bond less attractive, lowering its price. [See also **Seesaw effect**] Therefore, bondholders are subject to the risk of capital loss from such interest rate changes should the bonds have to be sold prior to maturity.

A longer term to maturity, all else being equal, increases the sensitivity of a bond's price to a given change in interest rates, as the discount rate change compounds over a longer time period. Similarly, a lower coupon rate also increases the sensitivity of the bond's price to market interest rate changes. This occurs since lower coupon bonds have most of their cash flow occurring further into the future, when the par value is paid.

Because of interest rate risk, investors will demand a larger risk premium for a bond whose price is especially sensitive to market interest rate changes. Hence, we

would expect higher yields to maturity for long-term bonds with low coupon rates than for short-term bonds with high coupon rates.

88. Interest Rate Structure

The concept that the interest rate or yield attached to any loan or security consists of the risk-free (or pure) rate of interest plus risk premiums for the security holder's exposure to default risk, inflation risk, call risk, and other forms of risk.

89. Interest Rates

The level and trend of interest rates play major roles in both financing and the investment decisions made by firms. Changes in interest rates may result in changes in a firm's bond and stock price as well as in the rates charged the firm by banks and other lenders. Such changes may affect the cost of financing enough to make an apparently profitable project turn unprofitable, or vice versa. The difference between long-term and short-term interest rates may influence a firm's decision to issue bonds or seek short-term financing.

Interest rates are the price of money. A borrower uses funds today, promising to repay them over time from future income. A saver forgoes current spending in order to store currency income in the expectation of earning a return that will increase the value of those savings over time. Thus, interest rates reflect the cost of moving income across time.

As with any price, interest rates rise and fall because of changes over time in demand and supply; in this case, the demand and supply of capital. There is not just one interest rate; there are a myriad of interest rates and, therefore, expected investment returns – from rates on short-term certificates of deposit at the bank, to rates offered on bonds issued by multinational corporations, to expected stock market returns. Interest rates or expected returns on these investments differ because of risk difference between them.

90. Interest Rate Swap

An interest rate swap is a financial transaction in which two borrowers exchange interest payments on a particular amount of principal with a specified maturity. The swap enables each party to alter the characteristics of the periodic interest payments that it makes or receives.

The exchange might involve swapping a fixed-rate payment for a variable payment or the payment of one type of floating rate for another. All swaps trade only interest payments made on underlying note values; no principal payments need to change hands with a simple interest rate swap.

The two primary parties to the swap are called counterparties. Usually, although not always, a financial institution serves as an intermediary between the counterparties. In the typical interest rate swap, the counterparty with the fixed-rate debt pays a premium over the rate the other counterparty initially paid on its

variable-rate debt. This premium is based upon factors such as the terms of the swap, the creditworthiness of the counterparties, and the conditions in the market for fixed-rate and variable-rate debt.

It is unusual for two companies to arrange an interest rate swap themselves. In most cases, intermediaries act as brokers, dealers, or principals to the transaction. As a broker or dealer, the intermediary serves to bring the counterparties together and collect an arrangement fee. However, in most swaps, the intermediary acts as a principal to both counterparties, assuming the credit risk in the event that one counterparty defaults. When the intermediary acts as the principal to a swap, its compensation is in the form of an arrangement fee and/or the spread between the terms of the two counterparties.

91. Interest Subsidy

A firm's deduction of the interest payments on its debt from its earnings before it calculates its tax bill under current tax law.

92. Interest-Rate Anticipation Swap

A rate anticipation swap is a bond trading strategy in which the trader exchanges the components of their bond portfolio in anticipation of expected interest rate movements.

93. Intermarket Spread Swap

Switching from one segment of the bond market to another (from Treasuries to corporates, for example).

94. Internal Audit

Routine examination of a bank's accounting records.

95. Internal Financing

Net income plus depreciation minus dividends, internal financing comes from internally generated cash flow such as retained earnings.

96. Internal Growth Rate

The internal growth rate measures how quickly a firm can increase its asset base over the next year without raising outside funds. It does not measure divisional growth or break down total growth into domestic or international components. More detailed analysis must be done to estimate the forces that determine growth rates. The internal growth rate gives a general, companywide value.

The internal growth rate is equal to the ratio of the expected increase in retained earnings (ΔRE) over the next year to the current total asset base (TA_0):

$$\frac{\Delta RE}{TA_0}$$

This also can be written as:

$$\text{Internal growth rate} = \frac{(RR)(ROA)}{1 - (RR)(ROA)}$$

RR is the firm's retention rate, and ROA is its return on assets. The internal growth rate divides the product of these values by one minus this product.

Most managers plan and think in terms of sales dollars rather than asset size, so it may help to relate the internal growth rate to sales growth.

We can roughly estimate a sales growth rate by recalling the total asset turnover ratio, which equals sales divided by total assets. If management can assume that this ratio will remain constant into the foreseeable future, the growth in sales will equal the internal growth rate.

The internal growth rate makes the restrictive assumption that the firm will pursue no outside sources of financing. Should the firm grow at its internal growth rate, its retained earnings account will continually rise (assuming profitable sales), while its dollar amount of debt outstanding will remain constant. Thus, the firm's debt-to-equity ratio declines over time, until debt falls below its correct proportion in management's ideal financing mix.

97. Internal Rate of Return

The internal rate of return (IRR) is a discounted cash flow concept and represents the discount rate at which the present value of a project's cash flows equal the project's cost. This implies that the IRR of a project is the discount rate that rate that sets the project's net present value (NPV) to zero. The internal rate of return can be found by solving for IRR in the following present value relationship:

$$NPV = \sum_{t=1}^N \frac{CF_t}{(1 + IRR)^t} - I = 0,$$

where:

I = initial outlay

CF_t = cash flow in period t

N = number of periods of project

If a project has a positive NPV when its cash flows are discounted at the required return, the project's IRR will exceed this required return. Graphically,

for any required rate of return below the IRR, the net present value of the project is positive. For any required rate of return greater than the IRR, the NPV is negative.

Thus, the IRR technique has a clearly defined and objective decision rule: accept all projects with IRRs that exceed their minimum required rate of return since those projects will increase shareholder wealth. If the IRR is less than the required return, then the project should be rejected, as it will reduce shareholder wealth. Both the NPV and IRR methods will always agree as to whether a project will increase or decrease shareholder wealth. Although the IRR and NPV are related, the meaning of the IRR is more complex than that of the NPV.

Unlike NPV, IRR does *not* measure the absolute dollar amount by which a project will change shareholder wealth. The IRR tells us nothing about the size of the change in shareholder wealth. Thus, we can say that IRR is a relative, not an absolute, measure of project attractiveness – one project's IRR may be higher, while its NPV is lower, than that of another project.

Thus, IRR satisfies only three of the four capital selection. It considers all cash flows, incorporates the time value of money, and it has an objective decision criterion. But it does not measure the size of the project's impact on shareholder wealth.

98. International Banking Act

A US law passed in 1978 to bring foreign banks operating in the United States under federal government regulation.

99. International Banking Facilities (IBFs)

A domestically based set of computerized accounts recording transactions of a US bank with its foreign customers and created by Federal Reserve Board regulation.

100. International Capital Asset Pricing Model

It is being used to test whether assets are best regarded as being traded in segmented (national) or integrated (international) markets, found some evidence that markets are integrated. It can be states as:

$$E(R_i^j) = R_f + \beta_{wi}^j [E(R_w) - R_f]$$

where:

$E(R_i^j)$ = expected rate of return on i th security (or portfolio) in country j

R_f = the risk-free rate of interest

$E(R_w)$ = expected rate or return on the world market portfolio

$\beta_{wi}^j = (\rho_{i,w}\sigma_i\sigma_w)/\sigma_w^2$ or the correlation coefficient between the rate of return on security i in country j and the world market, times the standard deviation of security i , times the standard deviation of the world market, divided by the variance of the world market portfolio. It is also the **international system risk** of country j .

101. International Finance

International finance, sometimes known as international macroeconomics, is the study of monetary interactions between two or more countries, focusing on areas such as foreign direct investment and currency exchange rates.

102. International Fisher Effect

The international Fisher effect shows (a) the relationship between the expected exchange rate change and the inflation rate differential, and (b) the relationship between the inflation rate differential and the interest rate differential. The equation is as follows:

$$\frac{E(S_1)}{S_0} = \frac{1 + h_{FC}}{1 + h_{US}} = \frac{1 + R_{FC}}{1 + R_{US}},$$

where S_0 and $E(S_1)$ are the current spot exchange rate and expected spot rate 1 year in the future respectively. R_{US} and R_{FC} are nominal interest rate in the United States and foreign country, respectively. Finally, h_{US} and h_{FC} are inflation rates for the United States and foreign country, respectively.

If the expected future spot rate, $E(S_1)$, is equal to the forward rate, F_1 , then the equation reduces to the interest rate parity relationship:

$$F_1 = S_0 \frac{(1 + R_{RC})}{(1 + R_{US})}$$

103. International Lending and Supervision Act

A federal law passed in 1983 requiring US banks to increase their capital and to pursue more prudent international loan policies.

104. International Monetary Market (IMM)

A division of Chicago mercantile exchange. The Eurodollar futures contracts and associated option are traded at IMM.

105. International Portfolio

An international portfolio is a selection of stocks and other assets that focuses on foreign markets rather than domestic ones. If well designed, an international

portfolio gives the investor exposure to emerging and developed markets and provides diversification.

106. International Stock Market Linkage

Linkage is the ability to buy a security on one financial exchange and sell that same security on another exchange. Certain depositary receipts, such as American Depositary Receipts (ADRs), allow for linkage, which means that an investor can purchase shares of a company on a foreign exchange, such as the Toronto Stock Exchange, and then sell those shares on a domestic exchange, such as the New York Stock Exchange.

107. International System Risk

[See **International asset pricing model**]

108. Interstate Operations

A trend toward financial institutions establishing facilities in more than one state.

109. Intertemporal Capital Asset Pricing Model

Allowing the investment opportunity set change over time, Merton (1973, pp. 363–406) develop the intertemporal capital asset pricing model which introduces a hedge portfolio function into the model.

110. In the Money

The owner of a put or call is not obligated to carry out the specified transaction but has the *option* of doing so. If the transaction is carried out, it is said to have been *exercised*. For example, if you hold a call option on a stock that is currently trading at a price higher than the exercise price, you may want to exercise the option to purchase stock at the exercise price and then immediately resell the stock at a profit. This call option is said to *be in the money*.

111. In-the-Money-Option

Either (a) a call option where the asset price is greater than the strike price or (b) a put option where the asset price is less than the strike price.

112. Intrinsic Value

For a call option, this is the greater of the excess of the asset price over the strike price and zero. For a put option, it is the greater of the excess of the strike price over the asset price and zero.

113. Intrinsic Value of an Option

Stock price minus exercise price, or the profit that could be attained by immediate exercise of an in-the-money option.

114. Inventory

A current asset, composed of raw materials to be used in production, work in process, and finished goods.

115. Inventory Conversion Period

The inventory conversion period is defined as inventory divided by cost of goods sold per day:

$$\text{Inventory conversion period} = \frac{\text{Inventory}}{\text{Cost of goods sold}/365 \text{ days}}$$

116. Inventory Loan

A secured short-term loan to purchase inventory. The three basic forms are blanket inventory lien, a trust receipt, and field warehouse financing.

117. Inventory Turnover Ratio

Ratio of annual sales to average inventory that measures how quickly inventory is produced and sold. [See **Asset management ratios**]

118. Inverted Market

A market where futures prices decrease with maturity.

119. Inverted Yield Curve

Yield curve with long-term rates below short-term rates.

120. Investable Balances

Ledge balances minus float minus required reserves against associated deposit balances.

121. Investment Asset

An asset held by at least some individuals for investment purposes.

122. Investment Bankers

Financial intermediaries who perform a variety of service, including aiding in the sale of securities, facilitating mergers and other corporate reorganizations, acting as brokers to both individual and institutional clients, and trading for their own accounts.

123. Investment Banking

Activity involving securities underwriting, making a market in securities, and arranging mergers and acquisitions.

124. Investment Company

Firm managing funds for investors. An investment company may manage several mutual funds.

125. Investment Decision

Investment decisions are made by investors and investment managers. Investors commonly perform investment analysis by making use of fundamental analysis, technical analysis, and gut feel. Investment decisions are often supported by decision tools. The portfolio theory is often applied to help the investor achieve a satisfactory return compared to the risk taken.

126. Investment Grade Bond

Debt that is rated BBB and above by Standard and Poor's or Baa and above by Moody's. Alternatively, it is lower-rated bonds are classified as speculative-grade or junk bonds.

127. Investment Horizons

Investment horizon is the term used to describe the total length of time that an investor expects to hold a security or a portfolio. Investment horizons can range from short-term, just a few days long, to much longer-term, potentially spanning decades. For example, a young professional with a 401(k) plan would have an investment horizon that would span decades. However, a corporation's treasury department might have an investment horizon that is only a few days long.

128. Investment Institutions

Financial intermediaries selling their customers securities and other financial assets in order to build up savings for retirement or for other uses.

129. Investment of Different Life

[See **Equivalent annual cost**]

130. Investment Opportunity Schedule

An investment opportunity schedule (IOS) is a chart or graph that relates the internal rate of return on individual projects to cumulative capital spending. To set up an investment opportunity schedule, the analyst first computes each project's internal rate of return (or modified internal rate of return). If mutually exclusive projects are part of the analysis, only the highest ranked projects go on to the next step.

After computing the individual modified internal rates of return (MIRR), the projects are ranked from highest to lowest by MIRR, keeping a tally of cumulative project spending.

131. Investment Portfolio

Set of securities chosen by an investor.

132. Investment Quality Bonds

Investment quality bonds have ratings of BBB by Standard and Poors (or Baa by Moodys) or higher. They are called investment quality as some institutional investors, such as pension funds and insurance companies, restrict themselves to investing only in these low-default risk issues.

133. Investment Trigger Price

The price of an investment project (or the price of the good to be produced) at which it is optimal to invest in the project.

134. Investments

An area within finance is the study of investments. Students of investments learn how to analyze the investor's stake in stocks, bonds, and other financial instruments. This analysis focuses on evaluating the cash flow from such financial assets to decide whether they represent attractive investments. As in the other fields of finance, the analyst also must plan how to manage the assets in an investment portfolio to meet future liabilities (such as college tuition, a new car or house, or retirement income).

135. Investor Sentiment

Market sentiment refers to the overall attitude of investors toward a particular security or financial market. It is the feeling or tone of a market, or its crowd psychology, as revealed through the activity and price movement of the securities

traded in that market. In broad terms, rising prices indicate bullish market sentiment, while falling prices indicate bearish market sentiment.

136. Invoice

Bill written by a seller of goods or services and submitted to the purchaser.

137. Invoicing Float

[See **Float**]

138. IO

Interest Only. A mortgage- backed security where the holder receives only interest cash flows on the underlying mortgage pool.

139. IPO Issuance and Performance

An initial public offering (IPO) refers to the process of offering shares of a private corporation to the public in a new stock issuance. Public share issuance allows a company to raise capital from public investors. The transition from a private to a public company can be an important time for private investors to fully realize gains from their investment as it typically includes share premiums for current private investors.

140. Irrelevance Result

The Miller and Modigliani (1958) theorem that a firm's capital structure is irrelevant to the firm's value when there are no taxes and other assumptions hold.

141. Irrevocable Letter of Credit

International trade often requires banker's acceptances, as well as even more formal arrangements. An exporter that requires even greater certainty of payment may request an irrevocable letter of credit. In this arrangement, the customer's bank sends the exporter a letter stating that it has established a line of credit for the customer with a particular US bank. The exporter then can collect payment from the US bank before making the delivery. The US bank forwards the appropriate documents to the customer's bank to receive reimbursement.

142. ISDA (Institutional Swap Dealers Association)

A committee sponsored by this organization was instrumental in drafting an industry standard under which securities dealers would trade swaps. Including in this draft of a master agreement by which institutions outlined their rights to net

multiple offsetting exposures which they might have to counterparty at the time in credit quality.

143. Iso-Expected Return Line

Line, drawn on a mapping of portfolio weights, which shows the combinations of weights all of which provide for a particular portfolio rate of return.

144. Iso-Variance Ellipse

Ellipse, drawn on a mapping of portfolio weights, which shows the combinations of weights all of which provide for a particular portfolio variance.

145. Issuer Exposure

The credit risk to the issuer of traded instruments (typically a bond, but also swaps, foreign exchange, etc.) Labeling credit spread volatility as either market or credit risk is a question of semantics. CreditMetrics addresses market price volatility as it is caused by changes in credit quality. [See **CreditMetrics**]

146. Itô Process

A stochastic process where the change in a variable during each short period of time of length (δt) has normal distribution. The mean and variance of the distribution are proportional to (δt) and are not necessarily constant.

1.10 J

1. January Effect

Market anomaly whereby stock prices throughout most of the world have a propensity to rise sharply during the initial part of the month of January.

2. Jensen's Inequality

If x is a random variable and $f(x)$ is convex, Jensen's inequality states that $E[f(x)] \geq f[E(x)]$. The inequality is reversed if $f(x)$ is concave.

3. Jensen's Measure

The alpha of an investment. It can be defined as:

$$JM = (\bar{R}_i - R_f) - \beta_i [\bar{R}_m - R_f]$$

where $\bar{R}_i$ is an average rate of returns for i th asset or portfolio; R_f = risk-free return; $\bar{R}_m$ = average market rates of return; and β_i is the beta coefficient for the i th asset.

4. Johnson Hedge Model

Developed within the framework of modern portfolio theory. The Johnson hedge model (1960, pp. 139–151) retains the traditional objective of risk minimization but defines risk as the variance of return on a two-asset hedge portfolio. As in the two-parameter world of Markowitz's (1959), the hedger is assumed to be infinitely risk averse (i.e., the investor desires zero variance). Moreover, with portfolio optimization, the risk-minimization objective defined as the variance of return on the combined spot and futures position, the Johnson hedge ratio is expressed in terms of expectations of variance and covariances for price changes in the spot and futures markets. The Johnson hedge model can be expressed in regression form as:

$$\Delta S_t = \alpha + H\Delta F_t + e_t$$

where:

ΔS_t = change in the spot price at time t

ΔF_t = change in the futures price at time t

α = constant

H = hedge ratio

e_t = residual term at time t

5. Joint Probabilities

In credit risk analysis, stand-alone obligors have some likelihood of each possible credit quality migration. Between two obligors there is some likelihood of each possible joint credit quality migration. The probabilities are commonly influenced by the correlation between the two obligors.

6. Joint Venture

A joint venture is a partial business combination. Two or more entities form a new corporation or partnership in order to jointly pursue a business venture. This provides an opportunity to combine resources in optimal proportions rather than in the fixed portfolio proportions dictated by a merger or a tender offer. The participants are partners rather than acquirer and target, and thus the formation of a joint venture does not cast one party as the aggressor, as in a merger or acquisition.

Common reasons for joint venture formation include facilitating technological transfer and developing market structures. International diversification also can give rise to joint ventures because some countries require local investment from any firm operating within their borders; others exempt firms with local participation from government regulations.

Joint ventures also can be used for undertaking certain massive projects. The development of Prudhoe Bay, Alaska, is one such example of a project joint venture.

7. Judgment

Legal ruling regarding the final payment of a court-determined transfer of assets.

8. Judgmental Credit Analysis

Subjective assessment of a borrower's ability and willingness to repay debts.

9. Jumbos

Jumbos are negotiable certificates of deposits (CDs) by thrifts which are large-denomination (\$100,000 or greater) time deposits with a minimum maturity of 7 days.

10. Jump-Diffusion Model

A process for an asset price in which the asset most of the time follows an Itô process but can also jump discretely, with occurrence of the jump controlled by a Poisson process.

11. Junior Liens

[See **Second mortgages**]

12. Junk Bonds

Noninvestment quality bonds are called junk bonds or **high-yield bonds** to reflect their higher risk and higher expected returns.

1.11 K

1. Kappa

Another name for *vega*. [See **Vega**]

2. Keogh Plan

A pension plan for the self-employed which allows them to make contributions and defer taxes until the funds are withdrawn.

3. Key Borrower

From bank-customer point of view, a key borrower refers to a company with significant business activities with the bank and receives the favorable lending rate from the bank.

4. Key-Person Insurance

Most banks require key-person insurance on the principal officers of the borrowing company to protect their loans. Because the repayment of a loan usually depends upon the managers of the firm running the company profitability, the death or disability of a key manager could jeopardize the safety of the loan. To avoid this uncertainty, the borrower buys a term insurance policy on the life of the key manager for the value of the loan. If he or she should die, the proceeds of the policy would be paid to the bank in settlement of the loan. Key-person insurance is useful in sole proprietorships as well as corporations.

5. Kite

Writing checks against uncollected deposits in the process of clearing through the banking system.

6. KMV-Merton Model

Merton model is known as the triggers to the development of many credit risk models. Of all the credit risk models developed, the KMV-Merton model is the most popular. KMV-Merton model is developed to provide probabilistic assessment of firm's likelihood to default.

7. Knock-In Option

An option in which there can only be a final payoff if, during a specified period of time, the price of the underlying asset has reached a specified level. This is one of the barrier options; it is attractive to some market participants because they are less expensive than the regular options.

8. Knock-Out Option

An option in which there can only be a final payoff if, during a specified period of time, the price of the underlying asset has *not* reached a specified level. This is one of the barrier options; it is attractive to some market participants because they are less expensive than the regular options.

9. Kolmogorov Backward Equation

A partial differential equation that is related to the Black-Scholes equation and that is satisfied by probability distributions for the underlying asset.

10. Kurtosis

Characterizes relative peakedness or flatness of a given distribution compared to a normal distribution. It is the fourth moment of a distribution. Since the unconditional normal distribution has a kurtosis of 3, excess kurtosis is defined as K_{x-3} . Sample kurtosis can be defined as:

$$K_{x-3} = \sum_{i=1}^n (X_i - \bar{X})^4 / n - 3$$

K_{x-3} can either be equal to, larger than, or smaller than 0.

[See **Leptokurtosis**]

1.12 L

1. L

The definition of the US money supply that includes M3 definition of money plus nonbank public holdings of US Savings Bonds, short-term Treasury securities, commercial paper, and bankers' acceptances, net of money market mutual funds' holdings of these same assets.

2. Ladder Option

If the barrier $L > K$ is reached over the life of the option, a ladder option at expiration pays $\max(0, L - K, S_T - K)$. If the barrier is not reached, the option pays $\max(0, S_T - K)$.

Ladder options are "more path-dependent" than barrier options. Normally, there are several pre-specified ladders or rungs in a ladder option. Whenever the underlying asset price reaches a pre-specified higher level in a series of prespecified rungs, the intrinsic value of the option is locked.

3. Ladder Strategy

When investing bonds, allocating roughly equivalent amounts (portions) to different maturities.

4. Lagged Reserve Accounting

System of reserve requirements based on deposits outstanding prior to the reserve maintenance period.

5. Lagging Indicators

[See **Business cycle**]

6. Lagrange Multiplier

In mathematical optimization, the method of Lagrange multipliers is a strategy for finding the local maxima and minima of a function subject to equality constraints (i.e., subject to the condition that one or more equations have to be satisfied exactly by the chosen values of the variables). The basic idea is to convert a constrained problem into a form such that the derivative test of an unconstrained problem can still be applied. The relationship between the gradient of the function and gradients of the constraints rather naturally leads to a reformulation of the original problem, known as the Lagrangian function. In portfolio analysis, we frequently use the Lagrange multiplier to determine the optimum waste in a portfolio.

7. Lambda

Another name for *vega*. [See **Vega**]

8. Latent Factors (Variables)

Latent variables may correspond to aspects of physical reality. These could in principle be measured, but may not be for practical reasons. In this situation, the term hidden variables is commonly used (reflecting the fact that the variables are meaningful, but not observable). Other latent variables correspond to abstract concepts, like categories, behavioral or mental states, or data structures. The terms hypothetical variables or hypothetical constructs may be used in these situations. In finance, latent factors have been frequently used in credit analysis.

9. Lattice

A binomial tree in which an up move followed by a down move leads to the same price as a down move followed by an up move. Also called a *recombining* binomial tree.

10. Law of One Price (LOP)

A commodity will cost the same regardless of what currency is used to purchase it. The LOP is also the guiding principle behind the Miller and Modigliani arbitrage argument.

11. LBO

[See **Leveraged buyout**]

12. Leading Economic Indicators

Economic series that tend to rise or fall in advance of the rest of the economy. [See **Business cycle**]

13. Leakage

Release of information to some persons before official public announcement.

14. Leaps

Long-term Equity Anticipation Securities. These are relatively long-term options on individual stocks or stock indices.

15. Lease

A contractual arrangement to grant the use of specific fixed assets for a specified time in exchange for payment, usually in the form of rent. An operating lease is generally a short-term cancelable arrangement, whereas a financial, or capital, lease is a long-term non-cancelable agreement.

16. Lease Accounting

Leases are contracts in which the property/asset owner allows another party to use the property/asset in exchange for something, usually money or other assets. The two most common types of leases in accounting are operating and financing (capital lease) leases. In order to differentiate between the two, one must consider how fully the risks and rewards associated with ownership of the asset have been transferred to the lessee from the lessor.

17. Lease Rate

The annualized payment required to borrow an asset, or equivalently, the annualized payment received in exchange for lending an asset.

18. Leasing Companies

Financial service firms that provide businesses and consumers access to equipment, motor vehicles, and other assets for a stipulated period of time at an agreed-upon leasing rate.

19. Ledger Cash

A firm's cash balance as reported in its financial statements. Also called book cash.

20. Ledger Balances

Dollar value of deposit balances appearing on a bank's books.

21. Legal Insolvency

Legal insolvency is a more serious financial problem than **technical insolvency**. [See also **Technical insolvency**] Legal insolvency exists when a firm's recorded assets amount to less than its recorded liabilities. This condition arises when successive losses create a deficit in the owners' equity account, rendering it incapable of supporting the firm's legal liabilities. The firm may be legally insolvent even when it is liquid and has plenty of cash to pay its current bills. Outsiders may not be aware of the insolvency as long as the liquidity of the firm enables it to meet its cash obligations. A protracted period of legal insolvency usually leads to bankruptcy.

22. Legal Lending Limit

The maximum amount that can be loaned to any one borrower or any group or related borrowers.

23. Legal Reserves

Deposits held at the Federal Reserve banks by depository institutions plus currency and coin held in the vaults of these institutions.

24. Lender Liability

Circumstances in which the courts have found lenders liable to their borrowers for fraud, deception, breached fiduciary activities, broken promises, and good faith negotiations.

25. Lending Portfolio

Investors invest in both market portfolio and risk-free asset. When they invest in risk-free asset, it means that they lend money to somebody else. Therefore, this kind of portfolio is called lending portfolio.

26. Leptokurtosis (Fat Tails)

The property of a statistical distribution to have more occurrences far away from the mean than would be predicted by a Normal distribution. Since a normal distribution has a kurtosis measure of 3, excess kurtosis (K_x) is defined as $K_x > 3$.

A credit portfolio distribution will typically be leptokurtotic given positive obligor correlations or coarse granularity in the size/number of exposures. This means that a downside confidence interval will be further away from the mean than would be expected given the standard deviation and skewness. [See **Kurtosis**]

27. Lessee

One that receives the use of assets under a lease. A long-term capital or financial lease obligates the lessee to make a series of fixed payments over time.

28. Lessor

One that conveys the use of assets under a lease. If the lessee fails to make payments as scheduled, the lessor, or the owner, can take possession of the leased asset.

29. Letter of Comment

A communication to the firm from the Securities and Exchange Commission that suggests changes to a registration statement. After the changes are made, the 20-day waiting period starts anew.

30. Letter of Credit

Time drafts increase the risk involved in foreign trade. To minimize this risk, an exporter may require the buyer to obtain a letter of credit (L/C) from a specified bank. Sometimes, buyers seek letters of credit themselves to obtain more favorable treatment by exporters. A letter of credit is a guarantee by the buyer's bank to honor the seller's drafts that are drawn on the bank, provided that the drafts comply with the terms specified in the letter of credit and are accompanied by the necessary documents.

A letter of credit affects a trade transaction in the following way. First, the buyer asks the bank to create the letter of credit in favor of the seller. Second, the bank creates the letter of credit and informs its foreign correspondent bank in the seller's country that it has done so. The correspondent bank in the seller's country then notifies the seller about the credit. Next, the seller ships the goods to the buyer and receives a **bill of lading (B/L)** from the shipper. [See also **Bill of lading**] Finally, the buyer sends the bill of lading to the shipper and receives the merchandise in return.

A letter of credit provides three important benefits to an importer:

1. An importer is safer if it deposits required prepayments with its own bank rather than with the seller in a foreign country. If the seller fails to ship the goods, it is relatively easy for the buyer to recover the deposit from its own bank.
2. If no prepayment is required, the buyer still can finance the purchase through its own bank at a relatively low cost.
3. The buyer can bargain for a lower price and better terms from the seller because it has substituted the bank's credit for its own. Since buyers who obtain letters of credit have eliminated most of the risk for the seller, they are justified in asking for lower prices and better terms.

A letter of credit also offers substantial advantages to the exporter. The exporter receives payment immediately after shipping the merchandise if the letter of credit specifies a sight draft. If the letter of credit calls for a time draft, the exporter receives a note from the bank (a **banker's acceptance**) rather than a note from the buyer; this bank note is virtually risk-free. [See also **Banker's acceptance**] Another advantage to a letter of credit comes from a reduction of the seller's risk of foreign exchange rate fluctuations due to the quick payment schedule.

There are three different types of letters of credit. (a) A *financial* letter of credit (also termed as stand-by letter of credit) is used to assure access to funding without the immediate need for funds and is triggered at the obligor's discretion. (b) A *project* letter of credit is secured by a specific asset or project income. (c) A *trade* letter of credit is typically triggered by a non-credit related (an infrequent) event. Item (c) is the abovementioned trade L/C.

31. Level-Coupon Bond

Bond with a stream of coupon payments that are the same throughout the life the bond. The coupon payments are equal to coupon rate times face value of a bond.

32. Leverage Effect

Leverage effect measures aim to quantify how much business risk a given company is currently experiencing. Business risk refers to the revenue variance that a business can expect to see, and how sensitive net income is to changes in revenues. Leverage effect measures aim to show how the business' fixed and variable costs can impact profitability when revenues change.

33. Leverage Ratio

Ratio of debt to total capitalization of a firm. [See **Capital structure ratios**]

34. Leveraged Buyout

A method of business combination is the leveraged buyout (LBO). In a leveraged buyout, the buyers borrow a major proportion of the purchase price, pledging the purchased assets as collateral for the loan. The buyers may be an outside group of investors, another company, or the manager of the firm or division that is being sold. Typically, the leverage arises from the payment of the purchase price to the seller (or alternatively, to a lender) using some of the actual earnings of the acquired firm. Once the assets are purchased, the cash flow from their operations is used to pay the principal and interest of the loan. In some cases, an LBO can be used to take a firm out of public ownership and into private ownership, in a technique called **going private**. Any kind of LBO can create an agency problem between the firm's managers and public shareholders, in that the managers usually have more and better information about the value of the firm than do the shareholders.

The LBO or merger method usually requires the target firm be either cash-rich (generate an abundant cash flow) or sell for less than the separate value of its assets. Additionally, forecasts of future cash flows for the target firm are necessary to estimate the riskiness of the deal over time.

35. Leveraged Equity

Stock in a firm that relies on financial leverage. Holders of leveraged equity face the benefits and cost of using debt. The required rates of return for a leveraged equity is higher than those of un-leveraged equity. [See **Modigliani and Miller (M&M) Proposition II**]

36. Leveraged Lease

In a leveraged lease, the lessor borrows money to purchase the asset and the leases out the asset. It is a tax-oriented leasing arrangement that involves one or more third-party lenders. This type of lease is often used in situations while large capital outlays are necessary for the purchase of asset.

37. Liabilities

Debts of the firm in the form of financial claims on a firm's assets. It can be classified as current liability and long-term liability.

38. Liability Management

The techniques used by banks to control the amount and composition of their borrowed funds by changing the interest rates they offer to reflect competition and the intensity and maturity of the bank's borrowing requirements.

39. Liability Management Theory

A theory that focuses in banks issuing liabilities to meet liquidity needs. Liquidity and liability management are closely related. One aspect of liquidity risk control is the buildup of a prudential level of liquid assets. Another aspect is the management of the deposit institution's (DI) liability structure to reduce the need for large amounts of liquid assets to meet liability withdrawals. However, excessive use of purchased funds in the liability structure can result in a liquidity crisis if investors lose confidence in the DI and refuse to roll over such funds.

40. Liability Sensitive

A bank is classified as liability sensitive if its GAP is negative.

41. LIBID

London Interbank Bid Rate. The rate bid by banks on Eurocurrency deposits (i.e., the rate at which a bank is willing to lend to other banks).

42. LIBOR

London Interbank Offer Rate. A measure of the borrowing rate for large international banks. The British Banker's Association determines LIBOR daily for different currencies by surveying at least eight banks, asking at what rate they could borrow, dropping the top and bottom quartiles of the responses, and computing an arithmetic average of the remaining quotes. Since LIBOR is an average, there may be no actual transactions at that rate. Confusingly, LIBOR is also sometimes referred to as a lending rate. This is because a bank serving as a market maker in the interbank market will offer to lend money at a high interest rate (LIBOR) and borrow money at a low interest rate (LIBID). (The difference between LIBOR and LIBID is the bid-ask spread in the interbank market.) A bank needing to borrow will thus pay LIBOR, and a bank with excess funds will receive LIBID. [See also **LIBID**]

43. LIBOR Curve

LIBOR zero-coupon interest rates as a function of maturity.

44. LIBOR-in-Arrears Swap

Swap where the interest paid on a date is determined by the interest rate observed on that date (not by the interest rate observed on the previous payment date).

45. Lien

Legal right granted by court to attach property until a legal claim is paid.

46. Life Insurance Companies

Financial service firms selling contracts to customers that promise to reduce the financial loss to an individual or family associated with death, disability, or old age.

47. LIFO

The last-in first-out accounting method of valuing inventories. In inflation period, the cost of inventory is higher than that calculated by the first-in-first-out method.

48. Limit Move

The maximum price move permitted by the exchange in a single trading session. There are both upper and lower limits.

49. Limit Order

An order that can be executed only at a specified price or one more favorable to the investor.

50. Limited Branching

Provisions that restrict branching to a geographic area smaller than an entire state.

51. Limited Liability

The fact that shareholders have no personal liability to the creditors of the corporation in the event of bankruptcy.

52. Limited Liability Company

A limited liability company (LLC) is one of two special forms of corporate organizations in the United States that allow dividends to escape double taxation. A limited liability company (LLC) organization form has been authorized by the laws of more than 35 states as of the end of 2005. Similar to a **Subchapter S corporation**, it offers owners limited liability and its income is taxed only once as personal income of the shareholder. [See also **Subchapter S corporation**] Unlike a Subchapter S corporation, however, an LLC can have an unlimited number of shareholders, including other corporations. The LLC can sell shares without completing the costly and time-consuming process of registering them with the SEC, which is a requirement for standard corporations that sell their securities to the public. The LLC structure has drawbacks in that, should an owner leave, all others must formally agree to continue the firm. Also, all of the LLC's owners must take active roles in managing the company. To protect partners from unlimited liability, some partnerships, including large accounting firms such as Coopers and Lybrand, Ernst and Young, and Price Waterhouse, have changed their organizational form to LLC.

53. Limited-Liability Instrument

A security, such as a call option, in which all the holder can lose is the initial amount put into it.

54. Limited Partnership

Form of business organization that permits the liability of some partners to be limited by the amount of cash contributed to the partnership. [See also **General partnership**]

55. Linear Programming Approach for Capital Rationing

The linear programming method can be applied to a multi-period capital rationing problem in one of two ways. The objective of the solution can be either (i) to maximize the total NPV from the investment in available projects or (ii) to maximize the present value of cash flow available for dividends.

56. Line of Credit

A line of credit is an agreement that specifies the maximum amount of unsecured credit the bank will extend to the firm at any time during the life of the agreement. In the past, banks gave lines of credit only to larger, more secure companies. This, too, appears to be changing, however, and some commercial banks now provide lines of credit to small, newly formed companies in which they see good growth potential.

In granting a line of credit, a bank is saying, in effect, "It looks as though your position is sufficiently sound to justify a loan, but when the time comes for you to start borrowing, we shall probably want to talk to you again to make sure that everything is going as expected." For example, a company that expects a rapid increase in sales may arrange a line of credit to finance increases in inventory and receivables. Before allowing the company to begin drawing on the line, however, the bank will want to verify that sales actually have increased. If the company has suffered a drop in sales, the bank is unlikely to allow it to use the line of credit to get out of the resulting financial crisis.

Of course, a line of credit has a cost to the borrower. When the loan actually is used, the borrower must pay interest on the funds borrowed. Even before actually accepting any funds, however, the borrower will probably incur a cost. Most banks require borrowers to keep a specified minimum compensating balance in exchange for being granted a line of credit. The compensating balance essentially compensates the bank for the service it provides. Instead of charging a fee for an additional interest rate, however, the bank obliges the borrower to keep an agreed-upon sum in its demand deposit account at all times. Since banks pay no interest on commercial demand deposits, they may then invest the compensating balance in marketable securities or lend them to another borrower; any return the bank earns on these funds is clear profit. In practice, the use of compensating balances has been dwindling. This is especially true for larger firms, which would rather pay fees than hold compensating balances.

57. Linear Programming Approach to Portfolio Analysis

Sharpe (1967, pp. 499–510) developed a simplified portfolio-analysis model designed to be formulated as a linear-programming problem. Sharpe approaches the problem of capturing the essence of mean-variance portfolio selection in a linear-programming formulation by:

1. Making a diagonal transformation of the variables that will convert the problem into a diagonal form
2. Using a piecewise linear approximation for each of the terms of variance, the LP that results from the use of market responsiveness as the risk measure and the imposition of an upper on investment in each security is

$$\text{MAX } P = \lambda \left[\sum_{i=1}^n x_i E(R_i) \right] - (1 - \lambda) \left[\sum_{i=1}^n x_i \beta_i \right]$$

Subject to:

$$\sum_{i=1}^n x_i = 1 \quad (0 \leq x_i \leq U)$$

where:

x_i = the fraction of the portfolio invested in security i

$E(R_i)$ = the expected returns on security i

β_i = the beta coefficient of security i

U = the maximum fraction of the portfolio that may be held in any one security

λ = a parameter reflecting the degree of risk aversion

58. Linear–Optimization Model

A linear-optimization model is a method of maximizing or minimizing an objective function that is subject to a number of linear constraints. The general form of the problem can be written as

$$\text{Max}(\alpha_1 x_1 + \alpha_2 x_2)$$

Subject to

$$-x_1 + 4x_2 \geq 0$$

$$x_1 + x_2 = 1$$

$$x_1, x_2 \geq 0$$

where α_1, α_2 are the percentages of a portfolio invested in securities 1 and 2, respectively. x_1 and x_2 are average rates of return for securities 1 and 2, respectively.

The problem is to maximize the return on the portfolio. As shown by the objective function, with the restriction stated above, the investment in security 2 should be at least 20% ($-x_1 + 4x_2 \geq 0$); and, as stated in ($x_1 + x_2 = 1$), the funds of the portfolio should be 100% invested. Further, the nonnegative conditions ($x_1, x_2 \geq 0$) preclude the short selling of either security 1 or 2.

59. Lintner's Observations

John Lintner's (1956) work suggested the dividend policy is related to a target level of dividends and the speed of adjustment of change in dividends. [See **Lintner's model**]

60. Lintner's Model

Lintner's model, sometimes referred to as the *partial-adjustment model*, assumes that firms adjust their dividend payouts slowly over time and provides another explanation for a firm's dividend policy. In Lintner's model, a firm is assumed to have a desired level of dividends that is based on its expected earnings. When earnings vary, the firm will adjust its dividend payment to reflect the new level of earnings. However, rather than doing so immediately, a firm will choose to spread (or partially adjust) these variations in earnings over a number of time periods

Lintner (1956, pp. 97–113) was the first to investigate the partial-adjustment model of dividend behavior. Using this model, Lintner demonstrated how dividend policy decisions can be made by using the following three steps:

Step 1: Compare last period's dividend with the desired level of dividends and adjust the deviation accordingly the next period.

Step 2: Assume the desirable dividend level is $D^*_t = PT_t$, where P is the long-run payout ratio for dividends and E_t is the earnings level for that period.

Step 3: Combining steps 1 and 2, we obtain a dividend decision model:

$$D_t = D_{t-1} + \delta D^*_t - D_{t-1} \quad (14)$$

Or

$$D_t = D_{t-1} + \delta(PE_t - D_{t-1}) \quad (15)$$

To solve for the variable δ , the partial adjustment coefficient, we use a regression model:

$$D_t = b_0 + b_1 E_t + b_2 D_{t-1} + e_t \quad (16)$$

where $b_1 = P\delta$ and $b_2 = (1 - \delta)$. From the estimated b_1 and b_2 , we can estimate δ and P as follows:

$$\delta = 1 - \hat{b}_2 \quad (17)$$

$$P = \hat{b}_1 / (1 - \hat{b}_2) \quad (18)$$

Here we use the desired dividend payment, D_t , as a function of earnings, whereas with ratio analysis the desired ratio is a function of the industry average.

From this model, we can conclude that firms set their dividend in accordance with their level of current earnings. We can also conclude that changes in dividends over time do not correspond exactly with changes in earnings in the immediate time period, but rather are spread out over several time periods.

Another explanation for the δ coefficient is that it is the average speed of adjustment. We can interpret the quantity $(1 - \delta)$ as a safety factor that management uses to avoid increasing the dividend payment to levels that cannot be maintained.

Equation 16 shows the changes in dividend levels between periods rather than the absolute levels themselves. This allows us to investigate changes in the firm's dividend policy. Of the 28 firms Lintner studied, 26 appeared to have and follow a predetermined target payout ratio, P . On the whole, most of these firms updated their dividend policies annually.

61. Liquid Yield Option Note

First issued in 1985 after its development by Merrill Lynch a LYON is a Liquid Yield Option Note. In less fancy terms, LYONs are zero coupon, convertible, callable, puttable bonds.

They work this way: prior to maturity, an investor can convert the LYON into a specified number of common shares. As the value of the zero coupon bond approaches par over time, the conversion price increases according to a schedule set in the indenture. On designated dates prior to maturity, an investor can put the bond to the issuer and receive specific prices that increase as the value of the zero coupon bond approaches par over time. Finally, the issuer can call the bonds and pay investors an indenture-specified price that rises over time as the bond value accrues to par.

62. Liquidating Dividend

Payment by a firm to its owners from capital rather than from earnings.

63. Liquidation

Bankruptcy law favors reorganization through **Chapter 12**, but if the firm cannot be preserved as a going concern, the law requires liquidation. [See also **Chapter 12**] Liquidation involves selling the firm's assets and distributing the proceeds to the creditors in order of the priority of their claims. Chapter 8 of the Bankruptcy Reform Act of 1978 deals with "straight" liquidation.

In determining whether or not to liquidate a firm, the law asks: Is the firm worth more dead or alive? In other words, is the net present value of the liquidate parts of an enterprise greater than the present value of the firm as a going concern? If the answer is *yes*, the firm's assets are sold and the creditors are paid off. If the answer is *no*, then Chapter 12 proceeding usually are followed.

Once the liquidation of assets has begun, it usually becomes painfully clear that few, if any, assets except cash bring the balance sheet values. Indeed, a significant

reduction in asset values is to be expected. Because of this, not all claims on these assets will be satisfied in full; no liquidation generates enough cash to cover all claims.

In this even, available cash must be allocated to the various claims according to a rule called the **absolute priority of claims**. [See also **Absolute priority of claims**]

64. Liquidation Value

Net amount that could be realized by selling the assets of a firm after paying the debt.

65. Liquidity

Refers to the ease and quickness of converting assets to cash. Also called marketability. Current assets have higher liquidity than fixed assets. There are two separate meanings:

- (a) At the enterprise level, the ability to meet current liabilities as they fall due; often measures as the ratio of current assets to current liabilities.
- (b) At the security level, the ability to trade in volume without directly moving the market price; often measured as bid/ask spread and daily turnover.

66. Liquidity Preference Hypothesis

The liquidity preference hypothesis argues that long-term rates typically are higher than short-term rates because longer-term securities are inherently riskier than shorter-term securities; thus, long-term interest rates should incorporate a risk premium over and above the rates predicted by the expectations hypothesis.

Long-term bonds appear to be riskier than short-term bonds for several reasons. First, long-term bonds have greater interest rate risk; their prices change by larger percentages than short-term bond prices for the same change in market interest rates. Second, long-term bonds expose investors to more uncertainty about future inflation and interest rates.

Combining this risk premium perspective with the expectations hypothesis explains the term structure behavior better than the expectations hypothesis alone. [See also **Expectations hypothesis**] The term structure typically should slope upward, presumably due to the liquidity preference-risk premium effect. The term structure may become downward sloping, however, with long-term rates below short-term rates, if substantial declines in future rates are expected.

67. Liquidity Preference Theory

A theory leading to the conclusion that forward interest rates are above expected future spot interest rates.

68. Liquidity Preference Theory of Interest Rates

An explanation of the level of and changes in interest rates that focuses upon the interaction of the supply of and demand for money.

69. Liquidity Premium

The premium included in longer-term interest rates to compensate investors for price risk associated with volatile interest rates. This premium is due to the belief that most investors find long-term securities to be riskier than short-term securities. This hypothesis is called as liquidity premium hypotheses.

70. Liquidity Ratios

Liquidity ratios measure the ability of a firm to meet its maturing financial obligations and recurring operating expenses. In general, these are short-term obligations, normally due within 1 year. Several ratios provide evidence of liquidity.

The **current ratio** is defined as current assets (cash, marketable securities, accounts receivable, inventories, and prepaid expenses) divided by current liabilities (typically, accounts payable and short-term bank loans).

Some current assets, such as inventories and prepaid expenses, may not be very liquid. To assess liquidity without these questionable items, another liquidity measure called the **quick**, or **acid-test, ratio** may be used. The numerator of the quick ratio includes only cash, short-term marketable securities, and accounts receivable. The quick ratio is computed as current assets minus inventories and prepaid expenses divided by current liabilities. These two ratios are written as follows:

$$\text{Current ratio} = \frac{\text{Current assets}}{\text{Current liabilities}}$$

$$\begin{aligned} &\text{Quick (Acid-test) ratio} \\ &= \frac{\text{Current assets} - \text{Inventory} - \text{Other current assets}}{\text{Current liabilities}} \end{aligned}$$

Higher values for the liquidity ratios do not always imply greater liquidity and safety for short-term creditors. The current ratio would increase from 1 year to the next if the firm undertook an inventory buildup in anticipation of consumer demand that never occurred. The quick ratio helps to control for this distortion. Still, an increase in accounts receivable could result from either a poor credit check system or slow customer payment on accounts; either scenario could deceive analysts into representing the firm as more liquid than it really was.

A firm is liquid if it has the ability to raise sufficient funds quickly. The statement of cash flows can provide additional insight into the financial flexibility of a company and supplement liquidity ratio analysis.

71. Liquidity Risk

The variation in net income and market value of bank equity caused by a bank's difficulty in obtaining immediately available funds, either by borrowing or selling assets. It also refers the risk that a sudden surge in liability withdrawals may leave a financial institution in a position of having to liquid assets in a very short period of time and at low prices.

72. Liquidity Shocks (Crisis)

A liquidity crisis is a financial situation characterized by a lack of cash or easily convertible-to-cash assets on hand across many businesses or financial institutions simultaneously. In a liquidity crisis, liquidity problems at individual institutions lead to an acute increase in demand and decrease in supply of liquidity, and the resulting lack of available liquidity can lead to widespread defaults and even bankruptcies.

73. Liquidity-Based CAPM

Including illiquidity betas to the CAPM model contribute the explanation power of systematic risks on asset returns. The assets' liquidity commonality with the market and the covariance between assets' illiquidity and market returns present the importance of these illiquidity betas as significant risk factors on asset returns. On the other hand, assets' return sensitivity to the market liquidity has a positive and significant effect on the asset returns, although it is expected to be negative according to the theory.

74. Load Fund

A mutual fund with a sales commission, or load.

75. Loan Amortization

Individuals often borrow funds through amortized loans, including car loans and home mortgages. Under loan amortization, a loan is repaid by making equal or annuity payments over time. Each payment pays interest and repays some of the principal. The PVIFA interest factors, which determine annuity payments, aid the analysis of amortized loans.

Interest is a tax-deductible expense for home mortgages and business loans. For tax purposes, it is important to determine how much of each loan payment covers interest and how much constitutes return of principal. A tool to assist this process is a loan amortization schedule, which offers a year-by-year (or period-by-period) summary of the beginning loan balance, the annuity payment, the interest paid, the principal repaid, and the ending balance. The interest paid always equals the beginning periodic balance multiplied by the periodic interest rate. The principal repaid is always the difference between the total payment and the interest paid. The

ending balance represents the outstanding principal; it is computed by subtracting the principal repaid from the beginning balance of the period.

76. Loan Commitment

Formal agreement between a bank and borrower to provide a fixed amount of credit for a specified period.

77. Loan Exposure

The face amount of any loan outstanding plus accrued interest plus. [See **Dirty price**]

78. Loan Option

A contract entitling a borrower to take out a loan at a guaranteed interest rate over a stipulated time period.

79. Loan Participation

Credit extended to a borrower in which members of a group of lenders, each provide a fraction of the total financing; typically arises because individual banks are limited in the amount of credit they can extend to a single customer.

80. Loan Syndication

An arrangement where several lenders make a loan jointly to a borrower.

81. Loanable Funds Theory of Interest Rates

The credit view of what determines the level of and changes in interest rates which focuses upon the interaction of the demand for and the supply of loanable funds (credit).

82. Loan-to-Value Ratio

The loan amount divided by the appraised value of the underlying collateral.

83. Locals

Individuals on the floor of an exchange who trade for their own count rather than for someone else.

84. Location Risk

This is one of the components of the basic risk. The hedger requires delivery of the futures contract in location Y, but the only futures contracts available are for

delivery in location X. Hence, the hedger cannot form a perfect hedge because of the transportation costs from X to Y; this may cause the basis to change.

85. Lock-In Options

A lock-in option is an option which allows its holder to settle the option payoff at a time before the contracted option maturity, but transactions take place only at the expiration date. There are European lock-in options and American lock-in options. Whereas the lock-in time is prespecified in a European lock-in option, it is not contracted *ex ante* but can be chosen by the option holder at any time until the payment date in an American lock-in option.

While European lock-in options are less costly than vanilla options because of smaller time values and delayed payment of option payoffs, American lock-in options permit an investor to fix the option payoff at a more favorable time than merely waiting until the option expires.

86. Lockbox System

The primary distinguishing feature of a lockbox system is that the firm pays the local bank to take on the administrative chores. Instead of customers mailing their payments to one of the company's offices, they send all payments directly to a post-office box. One or more times a day, the bank collects the checks from the box and deposits them for collection. Among the advantages of a lockbox arrangement is the potential reduction in mail float and a significant reduction in processing float. In some more sophisticated arrangements, banks capture daily invoice data on magnetic tape and forward this data to the company's central office, thereby reducing the burden on the firm's accounts receivable staff.

87. Lock-up Provisions

IPOs usually contain lock-up provisions that forbid investors (such as corporate officers and directors, or investors such as venture capitalists who own large amounts of the newly public firms' shares) from selling their shares until a certain time after the IPO. By law, insiders must retain their shares for 90 days after the IPO, although some prospectuses required them to hold the shares even longer. The main reason for the lock-up provision is to prevent insiders from selling what may turn out to be overpriced stock immediately after the offering. Insiders can sell their shares as part of the IPO, but this information must be disclosed in the prospectus. Such selling typically is discouraged by the investment bank, however, as insider selling at the IPO sends a bad signal to the market about the insiders' optimism for the firm's future.

88. Logical Analysis of Data (LAD)

Logical analysis of data (LAD) is a methodology developed since the late eighties, aimed at discovering hidden structural information in data sets. LAD was originally developed for analyzing binary data by using the theory of partially

defined Boolean functions. LAD is frequently used to analyze credit risk analysis in terms of logistic regression.

89. Logistic Regression

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (a form of binary regression). Logistic regression has been frequently used to predict the default probability of a firm.

90. Lognormal Distribution

A variable has a lognormal distribution when the logarithm of the variable has a normal distribution.

91. London Interbank Offered Rate (LIBOR)

Rate the most creditworthy banks charge one another for large loans of Eurodollars overnight in the London market.

92. Long

A position is long with respect to a price if the position profits from an increase in that price. An owner of a stock profits from an increase in the stock price and, hence, is long the stock. An owner of an option profits from an increase in volatility and, hence, is long volatility.

93. Long Call

Long call means an investor buys an existing call option in the market. In general, an investor will buy a long call when they expect a stock to have upside potential.

94. Long Forward

The party to a forward contract who has an obligation to buy the underlying asset.

95. Long Hedge

Protecting the future cost of a purchase by purchasing a futures contract to protect against changes in the price of an asset.

96. Long Position

The purchase of futures contract in anticipation of taking eventual delivery of the commodity (or financial instrument) or an expected increase in the underlying asset's price.

97. Long Put

Long put means an investor buys an existing put option from the market. Long put is used to protect the existing stock owned by an investor when that stock is deemed to have downside risk.

98. Long Run

A period of time in which all costs are variable. It is an economics concept instead of accounting concept.

99. Long Short-Term Memory (LSTM)

Long short-term memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behavior required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning. This kind of technique can be applied in the technical analysis of an investment decision.

100. Long Straddle

A straddle is a simultaneous position in both a call and a put on the same underlying asset. A long straddle involves purchasing both the call and the put. By combining these two seemingly opposing options an investor can get the best risk-return combination that each offers. For a long straddle position. The profit potential is unlimited on upside, limited on downside. The loss potential is limited to the cost of call and put premiums. The effect of time decay is negative. The market sentiment is bullish or bearish. Thus a long straddle is an effective strategy for someone expecting the volatility of the underlying asset to increase in the future.

101. Long Vertical Spread

A spread is a combination of any two or more of the same type of options (two calls or two puts, for instance) on same underlying asset. A vertical spread specifies that the options have the same maturity month. Finally, a long vertical spread designates a position for which one has bought a low-exercise-price call (or a low-exercise put) that both mature in the same month. A long vertical spread is also known as a bull spread because of the bullish market expectation of the investor who enters into it. The investor limits the profit potential in selling the high-exercise-price call (or put). It is a popular position when is expected that the market will more likely go up than down. The profit potential is limited up to the higher exercise price. The loss potential is limited down to the lower exercise price. The effect of time decay is mixed. And the market expectation is cautiously bullish.

102. Long-Term Debt

An obligation having a maturity of more than 1 year from the date it was issued. Also called funded debt.

103. Long-Term Debt Ratio

Long-term debt ratio is long-term debt divided by the sum of long-term debt and total equity. This ratio is used to analyze the capital structure of a company.

104. Long-Term Securities

Securities with maturities in excess of 1 year.

105. Lookback Call

See *Lookback option*.

106. Lookback Option

An option that, at maturity, pays off based on the maximum ($\bar{S}_T$) or minimum ($\underline{S}_T$) stock price over the life of the option. A lookback call has the payoff $S_t - \underline{S}_T$ and a lookback put has the payoff $\bar{S}_T - S_t$, where S_t is the sales price of the stock at time t .

107. Lookback Put

See *Lookback option*.

108. Loss Given Default (LGD)

The loss severity of individual loan. It should take into account any collateral or guarantees. Both LGD and the probability of default (PD) are needed for a two-dimensional internal rating system.

109. Loss Reserve

Both life and property and liability insure estimate expected future claims on the exciting policies. These estimates are called loss reserve.

110. Low Discrepancy Sequence

[See **Quasi-random sequence**]

111. Low-Grade Bond

Junk bond.

112. Lower-of-Cost or Market Value Method

One of the two methods used to report an investment in other company. This method is used if no evidence of significant control exists. These securities are handled in the same way as marketable security

[See **Equity method**]

1.13 M

1. M1

The narrowest definition of the US money supply consisting of currency outside the treasury Federal Reserve banks, and the vaults of banks, plus checking accounts and other checkable deposits held by the nonbank public.

2. M2

The definition of the US money supply that includes M1 plus savings and small-denomination (under \$100,000) time deposits, money market mutual funds not held by institutions money market deposit accounts (MMDAs), overnight Eurodollar deposits issued to US banks worldwide, and overnight and continuing-contract repurchased agreements issued by all commercial banks.

3. M3

The definition of the US money supply which includes M2 plus large denominations (\$100,000-plus) time deposits and term repurchase agreements issued by commercial banks and thrift institutions, term Eurodollars held by US residents at foreign branches of US banks worldwide and at all banking offices in the United Kingdom and Canada, and institutional owned balances in money market mutual funds.

4. Macaulay Duration

The percent change in a bond's price for a given percent change in one plus the bond's yield. This calculation can be interpreted as the weighted average life of the bond, with the weights being the percentage of the bond's value due to each payment. A key assumption to this duration is that the yield curve is flat and that when rate changes, the yield curve shifts in a parallel fashion. [See also **Duration and Modifies duration**]

5. Machine Learning

Machine learning is the use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data. Machine Learning Methods are used to make the system learn using methods like Supervised learning and Unsupervised Learning which are further classified in methods like Classification, Regression and Clustering. There are various applications (like image classification, Predictive analysis, Spam detection) that use these different machine learning methods. Most recently, finance researchers use machine learning to analyze the probability of credit card default and other financial analysis. It should be known that machine learning uses nonparametric instead of parametric statistics.

6. Macroforecasting

Forecasts of price movements of the general stock market as a whole.

7. Macrohedge

A hedge strategy designed to reduce risk associated with a bank's entire balance sheet position. In other words, a financial institution manager wishes to use futures or other derivative securities to hedge the entire balance sheet duration gap.

8. Make-Whole Clause

A provision which requires that the borrower make a payment to a lender after a loan is called or prepaid. The amount of the payment equals the net present value of the lost interest and principal payments.

9. Mail Float

Refers to the part of the collection and disbursement process where checks are trapped in the postal system. [See **Float**]

10. Maintain Margin

Due to the difficulty of calling our customers whose margin accounts have fallen in value for the date, a clearing house member firm usually will require that a sum of money be deposited as an initiation of any future position. This additional sum is called maintain margin.

11. Maintenance Margin

Due to the difficulty of calling all customers whose margin accounts have fallen in value for the day, a clearing member firm usually will require that a sum of money

be deposited at the initiation of any futures position. This additional sum is called a maintenance margin. In most situations, the original margin requirement may be established with a risk-free, interest-bearing security such as a T-bill. However, the maintenance margin, which must be in cash, is adjusted for daily changes in the contract value. [See also **Marking to market**]

12. Make a Market

The obligation of a specialist to offer to buy and sell shares of assigned stocks. It is assumed that this makes the market liquid because the specialist assumes the role of a buyer for investors if they wish to sell and a seller if they wish to buy.

13. Making Delivery

Refers to the seller's actually turning over to the buyer the asset agreed upon in a forward contract.

14. Managed Float

An international monetary and payments system in which the value of any currency is determined by the demand and supply forces in the marketplace, but governments intervene on occasion in effort to stabilize the value of their own currencies.

15. Managed Floating Currencies Standard

System of currency valuation used today in which each nation chooses its own currency exchange rate standard or policy.

16. Management Risk

The variability of return caused by bad management decisions; this is usually a part of the unsystematic risk of a stock, although it can affect the amount of systematic risk.

17. MAPB System

The moving-average with a percentage price band system (MAPB system) belongs to a technical family derived from moving averages. Moving averages come in many forms—that is, simple moving averages, exponentially weighted, linearly weighted, and so on. The MAPB system employs a simple moving average with a band based on a percentage of price centered around it. A signal to initiate a position occurs whenever the closing price breaks outside the band. A signal to exit a position occurs when the price recrosses the moving average. The band creates a neutral zone in which the trader is neither long nor short.

18. Margin

A cash amount of funds that must be deposited with the broker for each futures contract as a guarantee of its fulfillment.

19. Margin Call

A demand for additional cash funds for each futures position held because of an adverse price movement.

20. Margin Requirement

Whenever someone enters into a contract position in the futures market, a security deposit, commonly called a margin requirement, must be paid. While the futures margin may seem to be a partial payment for the security on which the futures contract is based, it only represents security to cover any losses that may result from adverse price movements.

The minimum margin requirements set by the exchange must be collected by the clearing member firms (members of the exchange involved in the clearinghouse operations) when their customers take positions in the market. In turn, the clearing member firms must deposit a fixed portion of these margins with the clearinghouse. [See also **Marking to market**]

21. Marginal Cost of Funds

The incremental cost of additional funds to finance firm operations. Banks generally use federal funds or negotiable CD rates as marginal cost of funds.

22. Marginal Standard Deviation

Impact of a given asset on the total portfolio standard deviation.

23. Marginal Statistic

A statistic for a particular asset which is the difference between that statistic for the entire portfolio and that for the portfolio not including the asset.

24. Marginal Tax Rate

The marginal tax rate represents the proportion of each additional dollar of income that the government claims as taxes. The firm's marginal tax rate is important in financial decision making. Financial decisions hinge not on the **average tax rate** the firm has paid, but on the tax rate that applies to the additional income to be generated by a project. [See also **Average tax rate**]

One of the marginal costs that businesses must consider is the marginal tax that is paid should expansions or new projects be undertaken.

25. Marked to Market

Describes the daily settlement of obligation on futures positions. [See **Marking to market**]

26. Market Anomalies

The idea of an efficient market is very important to the study of security analysis and portfolio management. If information is fully reflected in security prices, the market is efficient and it is not worthwhile to pay for information that is already impounded in security prices. The evidence seems to indicate that markets are efficient with respect to most types of information. However, there appears to be certain types of information associated with irregularities in the financial markets. Such irregularities are called market anomalies. Three of the most heavily researched anomalies are the P/E effect, the size effect and the January effect.

27. Market Broadening

A Tendency for financial service markets to expand geographically over time due to advances in technology and increased customer mobility.

28. Market Capitalization

Price per share of stock multiplied by the number of shares outstanding. It is total market value of equity.

29. Market Capitalization Rate

The market-consensus estimate of the appropriate discount rate for a firm's cash flows.

30. Market Clearing

Total demand for loans by borrowers equals total supply of loans from lender. The market clears at the equilibrium rate of interest.

31. Market Conversion Price

[See **Convertible bond**]

32. Market Corner

Owning a large percentage of the available supply of an asset or commodity that is required for delivery under the terms of a derivatives contract.

33. Market Exposure

For market-driven instruments, there is an amount at risk to default only when the contract is in-the-money (i.e., when the replacement cost of the contract exceeds the original value). This exposure/uncertainty is captured by calculating the netted mean and standard deviation of exposure(s).

34. Market Interest Rate, Bond

[See **Yield to maturity**]

35. Market Model

A one-factor model for returns where the index that is used for the factor is an index of the returns on the whole market. It can be defined as:

$$R_{it} = \alpha_i + \beta_i R_{mt} + e_{it},$$

where R_{it} and R_{mt} are rate of return of i th security and market rate of return in period t , respectively. [See **Beta coefficient** and **Scatter diagram**]

36. Market or Systematic Risk, Firm-Specific Risk

Market risk is risk attributable to common macroeconomic factors. Firm-specific risk reflects risk peculiar to an individual firm that is independent of market risk.

37. Market Order

A buy or sell order to be executed immediately at current market prices.

38. Market Portfolio

The market portfolio is comprised of all risky assets weighted in proportion to their market value. As such, the market portfolio is a completely diversified portfolio; it has no unsystematic risk. The returns on this portfolio will show only the effects of marketwide or systematic risk. Investors who desire complete diversification and who want to eliminate unsystematic risk will want to hold the market portfolio.

Increase and decreases in an asset's value over time will reflect its exposure to both systematic and unsystematic risk factors. One way to measure the systematic risk of an asset or portfolio is to compare its returns over time with those of the market portfolio. Changes in the value of an asset relative to that of the market portfolio will reflect the asset's exposure to systematic risk factors.

39. Market Price

The current amount at which a security is trading in a market. For example, IBM price per share is \$95. This is a market price.

40. Market Price of Risk

A measure of the extra return, or risk premium, that investors demand to bear risk. The reward-to-risk ratio of the market portfolio. [See **Sharpe ratio**]

41. Market Risk

Systematic risk. This term emphasizes the fact that systematic risk influences to some extent all assets in the market. [See **Systematic risk**]

42. Market Segmentation Hypothesis

The market segmentation hypothesis explains the same phenomenon in terms of differences in supply and demand between segments of the capital markets. Some participants, such as banks, mainly borrow and lend short maturity securities. Others, such as pension funds, are major participants in the long-term portion of the yield curve. If more funds are available to borrow relative to demand in the short-term market than in the long-term market, short-term interest rates will be lower and long-term rates will be higher than predicted by both the expectations and liquidity preference hypothesis. [See also **Expectations hypothesis** and **Liquidity preference hypothesis**]

The drawback to this perspective is that it does not explain very well the usual upward slope of the term structure, nor does it provide a good explanation for the levels of intermediate-term rates. In addition, the financial markets are not strictly segmented; many institutions issue and purchase both short-term and long-term securities.

43. Market Segmentation Theory

The theory that long- and short-maturity bonds are traded in essentially distinct or segmented markets and that prices in one market do not affect those in the other. [See **Market segmentation hypothesis**]

44. Market Stabilization

During the aftermarket, the managing investment bank tries to prevent any significant declines in the price of the issuer's shares; hence, this function by investment banks is sometimes also called market stabilization.

Investment banks do not want to be known for bringing firms public at excessive offering prices, nor do they want to be known for handling IPOs of poor-quality issuers. To help show the market that the bank will stand behind its IPOs, it risks its own money to support the firm by repurchasing any and all shares offered to it at the offering price. This effectively places a floor under the firm's stock price.

The investment bank acts as a signal to market investors. When a highly reputable investment bank places its own capital at risk to underwrite securities, the investing public can have a greater degree of confidence regarding the quality of the issue. If an investment bank is willing to sell shares on a commission basis only, that is a signal of a low-quality, high-risk offering.

45. Market Timer

An investor who speculates on broad market moves rather than on specific securities.

46. Market Timing

Asset allocation in which the investment in the market is increased if one forecasts that the market will outperform T-bills.

47. Market Value

The price at which willing buyers and sellers trade a firm's assets. In general, the market value is different from book value.

48. Market Value Added

A measure to identify successful firms that is growing in popularity is market value added (MVA). MVA measures the value created by the firm's managers; it equals the market value of the firm's stocks and debts minus the amount of money investors paid to the firm (their book value) when these securities were first issued. That is, market value added (MVA) equals:

$$\begin{aligned} &\text{Market value of stock} + \text{Market value of debt} - \text{Book value of stock} \\ &\quad - \text{Book value of debt} \end{aligned}$$

49. Market Value Ratios

A firm's profitability, risk, quality of management, and many other factors are reflected in its stock and security prices. Hence, market value ratios indicate the market's assessment of the value of the firm's securities.

The **price/earnings (P/E) ratio** is simply the market price of the firm's common stock divided by its annual earnings per share. Sometimes called the *earnings multiple*, the P/E ratio shows how much investors are willing to pay for each dollar of the firm's earnings per share. Earnings per share comes from the income statement, so it is sensitive to the many factors that affect the construction of an income statement, from the choice of GAAP to management decisions regarding the use of debt to finance assets. The price/earnings ratio is stated as:

$$P/E = \frac{\text{Market price per share}}{\text{Earnings per share}}$$

Stock prices are determined from the actions of informed buyers and sellers in an impersonal market. Stock prices reflect much of the known information about a company and are fairly good indicators of a company's true value. Although earnings per share cannot reflect the value of patents or assets, the quality of the firm's management, or its risk, stock prices can and do reflect all of these factors. Comparing a firm's P/E to that of the stock market as a whole, or with the firm's competitors, indicates the market's perception of the true value of the company.

While the P/E ratio measures the market's valuation of the firm relative to the income statement value for per-share earnings, the **price-to-book-value ratio** measures the market's valuation relative to balance sheet equity. The book value of equity is simply the difference between the book values of assets and liabilities appearing on the balance sheet. The price-to-book-value ratio is the market price per share divided by the book value of equity per share. A higher ratio suggests that investors are more optimistic about the market value of a firm's assets, its intangible assets, and the ability of its managers. The price-to-book-value ratio is stated as:

$$\text{Price-to-book-value ratio} = \frac{\text{Market price per share}}{\text{Book value per share}}$$

Market value indicators reflect the market's perception of the true worth of a firm's *future prospects*. As such, market perceptions of a firm's value are important to the financial analyst. However, the market may not be perfect; investors may become overly optimistic or pessimistic about a firm. The fact that a firm presently has a higher P/E or price-to-book-value ratio than its competition does not automatically imply that the firm is better managed or really deserves its higher valuation. Some firms may have low market value ratios because they truly deserve them; other firms may suffer from extreme and undeserved pessimism on the part of the market. High market value ratios can be similarly deceptive. The analyst must determine whether a firm deserves its market value ratios or not.

50. Market-Based Beta Forecasts

Market-based beta forecasts are based upon market information alone. Historical betas of firms are used as a proxy for their futures betas. This implies that the unadjusted sample beta, $\hat{\beta}_t$ is equal to the population value of future beta:

$$\beta_{t+1} = \hat{\beta}_t$$

alternatively, there may be a systematic relationship between the estimated betas for the first period and those of the second period, as shown by Blume (1971, pp. 1–10):

$$\hat{\beta}_{i,t+1} = a_0 + a_1 \hat{\beta}_{i,t}$$

in which $\hat{\beta}_{i,t+1}$ and $\hat{\beta}_{i,t}$ estimated beta for the i th firm in period $t + 1$ and t , respectively.

51. Market-Book Ratio

Market price of a share divided by book value per share. [See **Tobin's Q**]

52. Market-Driven Instruments

Derivative instruments that are subject to counterparty default (e.g., swaps, forwards, options). The distinguishing feature of these types of credit exposures is that their amount is only the net replacement cost (the amount the position is in-the-money- rather than a full notional amount).

53. Market-Maker

A trader in an asset, commodity, or derivative who simultaneously offers to buy at one price (the bid price) or to sell at a higher price (the offer price), thereby “making a market.”

54. Mark-to-Market

The daily adjustment of a futures trading account to reflect profits or losses due to daily changes in the value of the futures contract. [See **Marking to market**]

55. Market-Value-Weighted Index

An index of a group of securities computed by calculating a weighted average of the returns of each security in the index, with weights proportional to outstanding market value. [See **S&P index**]

56. Marketability

Refers to the ease and quickness of converting an asset to cash. Also called liquidity.

57. Marketability Risk

The variability of return caused by the commissions and price concessions associated with selling an illiquid asset. It is also called liquidity risk. Marketability is made up of two components: (1) the volume of securities that can be bought or sold in a short period of time without adversely affecting the price, and (2) the amount of time necessary to complete the sale of a given number of securities. Other things being equal, the less marketability a security, the lower its price or the higher its yield.

58. Marketable Securities

Some current assets might be called *near cash* or *cash equivalents*. These are marketable securities. They are marketable because they can be readily converted into cash at any time without disrupting the normal routine of business operations. This feature makes these securities almost as liquid as cash, so cash and marketable securities often are combined into a single line item in financial reports and working capital analysis.

59. Marketed Claims

Claims that can be bought and sold in financial markets, such as those of stockholders and bondholders.

60. Market-to-Book (M/B) Ratio

Market price per share of common stock divided by book value per share. It can be used as approximated measure of Tobin Q.

61. Mark-to-Market Swap

Reduces default risk by allowing the fixed rate to be reset when fixed and floating rates diverge substantially after the beginning of the swap.

62. Marking to Market

At the end of each trading day, every futures-trading account is incremented or reduced by the corresponding increase or decrease in the value of all open interest positions. This daily adjustment procedure is applied to the margin deposit and is

called marking to market. For example, if an investor is long on a yen futures contract and by the end of the day its market value has fallen \$1,000, he or she would be asked to add an additional \$1,000 to the margin account. Why? Because the investor is responsible for its initial value. For example, if a futures contract is executed at \$10,000 with an initial margin of \$1,000 and the value of the position goes down \$1,000, to \$9,000, the buyer would be required to put in an additional margin of \$1,000 because the investor is responsible for paying \$10,000 for the contract. If the investor is unable to comply or refuses to do so, the clearing member firm that he or she trades through would automatically close out the position. On the other hand, if the contract's value was up \$1,000 for the day, the investor might immediately withdraw the profit if he or she so desired. The procedure of marking to market implies that all potential profits and losses are realized immediately.

63. Markov Model

A Markov Model is a stochastic model which models temporal or sequential data, that is, data that are ordered. It provides a way to model the dependencies of current information with previous information. It is composed of states, transition scheme between states, and emission of outputs (discrete or continuous). Goals that can be accomplished by using Markov models include learning statistics of sequential data, predictions or estimations, and pattern recognition. The objective function of the Markov model can be either maximize profit or minimize risk.

64. Markov Process

A stochastic process where the behavior of the variable over a short period of time depends solely on the value of the variable at the beginning of the period, not on its past history. Alternatively it is a finite set of "states" and whose next progression is determinable solely by the current state. A transition matrix model is an example of a Markov process.

65. Markowitz Model

Professor Markowitz (1952) published a paper entitled "Portfolio select model." In this paper, he shows how to create a frontier of investment portfolios such that each of them had the greatest possible expected return, given their level of risk.

66. Markowitz Modern Portfolio Theory

Modern portfolio theory (MPT) is a theory on how risk-averse investors can construct portfolios to maximize expected return based on a given level of market risk. Harry Markowitz pioneered this theory in his paper "Portfolio Selection," which was published in the *Journal of Finance* in 1952.

67. Martingale

A zero-drift stochastic process.

68. Master Note

A borrowing arrangement between a corporation issuing commercial paper and an institution buying the paper in which the buying institution agrees to accept new paper each day up to a specified maximum amount.

69. Maturity

The date at which the principal of a note, draft, or bond becomes due and payable.

70. Maturity Date

The date on which the last payment on a bond is due.

71. Maturity Gap

It is the difference between the weighted-average maturity of financial institution's asset and liability. The maturity model with a portfolio of assets and liabilities is:

$$\text{Maturity gap} = M_A - M_L,$$

where M_A and M_L represent the weighted average maturity of FI's asset and liability, respectively.

72. Maturity Premium

When the default risk on a class of securities is virtually zero, the risk premium represents a maturity premium that reflects uncertainty about inflation and changes in interest rates over a longer time horizon.

73. Maximum Expected Utility Hedge Ratio

In order to make the hedge ratio consistent with the expected utility maximization principle, we need to derive the hedge ratio that maximizes the expected utility. However, in order to maximize the expected utility we need to assume a specific utility function. For example, Cecchetti et al. (1988) derive the hedge ratio that maximizes the expected utility where the utility function is assumed to be the logarithm of terminal wealth as follows:

$$\int \int_{R_s R_f} \log [1 + R_s - hR_f] f(R_s, R_f) dR_s dR_f$$

where R_f and R_s are the futures rate of return and the spot rate of return, respectively, the density function $f(R_s, R_f)$ is assumed to be bivariate normal. Specifically, they derive the optimal hedge ratio that maximizes the expected utility function. A third-order linear bivariate ARCH model is used to get the conditional variance and covariance matrix, and a numerical procedure is used to maximize the objective function with respect to the hedge ratio.

74. Maximum Likelihood Method

A method for choosing the values of parameters by maximizing the probability of a set of observations occurring.

75. MBS

Mortgage-backed security; a security that evidences an undivided interest in the ownership of a pool of mortgages.

76. Mean Reversion

The statistical tendency in a time series to gravitate back toward a long-term historical level. This is on a much longer scale than another similar measure, called autocorrelation; and these two behaviors are mathematically independent of one another. For example, the tendency of a market variable (such as an interest rate) to revert back to some long-run average level.

77. Mean-Variance Analysis

Evaluation of risky prospects based on the expected value and variance of possible outcomes. [See **Capital market line**]

78. Mean-Variance Capital Asset Pricing

Mean-variance analysis leads directly to the capital asset pricing model or CAPM. The CAPM is a one-period equilibrium model that provides many important insights to the problem of asset pricing. The language/jargon associated with the CAPM has become ubiquitous in finance. Asset pricing models can include skewness as an additional factor. Therefore, the mean-variance capital asset pricing does not consider the skewness effect of the rate of return.

79. Mean-Variance Criterion

The selection of portfolios based on the means and variances of their returns. The choice of the higher expected return portfolio for a given level of variance or the lower variance portfolio for a given expected return.

80. Measure

Sometimes also called a probability measure; it defines the market price of risk.

81. Measurement Error

Errors in measuring an explanatory variable in a regression that leads to biases in estimated parameters.

82. Median

The median is defined as the outcome value that exceeds the outcome value for half the population and is exceeded by the other half. Whereas the expected rate of return is a weighted average of the outcomes, the weights being the probabilities, the median is based on the rank order of the outcomes and takes into account only the order of the outcome values.

83. Member Banks

Banks that have joined the Federal Reserve System, consisting of all federally chartered (national) banks and any state chartered US banks that meet the Federal Reserve's requirements for membership.

84. Membership or Seat on an Exchange

A limited number of exchange positions that enable the holder to trade for the holder's own accounts and charge clients for the execution of trades for their accounts.

85. Merchandise Trade Balance

The difference between the value of a nation's exports of goods and the value of its imports of goods.

86. Merger

Assuming there are originally two firms, Firm A and Firm B. In one possible business combination, only Firm B survives. This type of combination is known as a

merger and Firm B is called the *acquiring firm* while Firm A is called the *acquired* or *target firm*. [See also **Acquisition**]

Many researchers in economics and finance have advanced theories to justify firms' pursuit of mergers. The most recent efforts by academicians in these fields have begun to integrate individual theoretical rationales.

Among the myriad reasons that have been proposed to explain merger activity, the more prominent ones are:

1. Economies of scale
2. Market power and market share
3. Diversification
4. Tax and surplus funds motives
5. Undervalued assets
6. Agency problems

87. Merrill Lynch Beta

Merrill Lynch uses the mean-reversion process to adjust beta toward one. In other words, the Merrill Lynch beta is the weighted-average of the actual beta and one.

88. Merton Distance Model

One of the most effective methods for rating credit risk is built on the Merton Distance to Default model, also known as simply the Merton Model.

89. Microeconomic Risk

The risk that is diversified away as assets are added to a portfolio is the firm- and industry-specific risk, or the "microeconomic" risk. A well-diversified portfolio can reduce the effects of firm- or industry-specific events – such as strikes, technological advances, and entry and exit of competitors – to almost zero.

90. Microhedge

A hedge strategy designed to reduce the risk associated with a specific transaction. In other words, a financial institution employs a futures or forward contract to hedge a particular asset or liability risk.

91. Migration

Credit quality migration describes the possibility that a firm or obligor with some credit rating today may move to (or "migrate") to potentially any other credit rating – or perhaps default – by the risk horizon. [See **Migration matrix**]

92. Migration Analysis

The technique of estimating the likelihood of credit quality migration. [See **Transition matrix**]

93. Miller-Orr Model

The Miller-Orr model (1966, pp. 413–435) for cash management improves on Baumol's economic order quantity model (EOQ) methodology in significant ways. Miller and Orr start with the assumption that the firm has only two forms of assets: cash and marketable securities. The model allows for cash balance movement in both positive and negative directions and it can state the optimal cash balance as a range of values, rather than a single-point estimate. This makes the model especially useful for firms with unpredictable day-to-day cash inflows and outflows.

While the Miller-Orr model is an improvement over the **EOQ model**, it too makes some assumptions. [See also **Baumol's economic order quantity model**] The most important is the assumption that cash flows are random, which in many cases is not completely valid. Under certain circumstances and at particular times of the year, consecutive periods' cash flows may be dependent upon one another, the volatility of net cash flows may sharply increase, or cash balances may demonstrate a definite trend. The frequency and extent of these events will affect the Miller-Orr model's effectiveness. Actual tests using daily cash flow for various firms indicate that the model minimizes cash holding costs as well as or better than the intuitive decisions of these firms' financial managers. However, others studies have shown that simple rules of thumb have performed just as well. Still, the Miller-Orr model is valuable because of the insight it offers concerning the forces that influence a firm's optimal cash balance.

94. Minimum Generalized Semivariance Hedge Ratio

In recent years, a new approach for determining the hedge ratio has been suggested by Chen et al. (2001). This new approach is based on the relationship between the generalized semivariance (GSV) and expected utility. In this case the optimal hedge ratio is obtained by minimizing the GSV, presented as follows:

$$V_{\delta,\alpha}(R_h) = \int_{-\infty}^{\delta} (\delta - R_h)^{\alpha} dG(R_h), \alpha > 0$$

where $G(R_h)$ is the probability distribution function of the return on the hedged portfolio R_h . The parameters δ and α (which are both real numbers) represent the target return and risk aversion, respectively. The risk is defined in such a way that the investors consider only the returns below the target return (δ) to be risky. It can be shown that $\alpha < 1$ represents a risk-seeking investor and $\alpha > 1$ represents a risk-averse investor.

The GSV, due to its emphasis on the returns below the target return, is consistent with the risk perceived by managers. Lien and Tse (1998) show that the GSV hedge ratio, which is obtained by minimizing the GSV, would be the same as the MV hedge ratio if the futures and spot returns are jointly normally distributed and if the futures price follows a pure martingale process.

95. Minimum Mean Extended-Gini Coefficient Hedge Ratio

This approach of deriving the optimal hedge ratio is consistent with the concept of stochastic dominance and involves the use of the mean extended-Gini (MEG) coefficient. It minimizes the MEG coefficient defined as follows:

$$\Gamma_{\nu}(R_h) = -\nu \text{Cov}\left(R_h, (1 - G(R_h))^{\nu-1}\right)$$

where G is the cumulative probability distribution in terms of the rate of return of a hedge portfolio and ν is the risk aversion parameter. Note that $0 \leq \nu < 1$ implies risk seekers, $\nu = 1$ implies risk-neutral investors, and $\nu > 1$ implies risk-averse investors. Shalit (1995) has shown that if the futures and spot returns are jointly normally distributed, then the minimum-MEG hedge ratio would be the same as the MV hedge ratio.

96. Minimum Value-at-Risk Hedge Ratio

This hedge ratio is obtained by minimizing the Value-at-Risk of the hedged portfolio. Specifically, the hedge ratio h is derived by minimizing the following Value-at-Risk of the hedged portfolio over a given time period.

$$\text{VaR}(R_h) = Z_{\alpha} \sigma_h \sqrt{\tau} - E[R_h] \tau$$

where R_h is the rate of return for the hedge portfolio and σ_h is the standard deviation of R_h .

The resulting optimal hedge ratio, which Hung et al. (2006) refer to as zero-VaR hedge ratio, is given by:

$$h^{\text{VaR}} = \rho \frac{\sigma_s}{\sigma_f} - E[R_f] \frac{\sigma_s}{\sigma_f} \sqrt{\frac{1 - \rho^2}{Z_{\alpha}^2 \sigma_f^2 - E[R_f]^2}}$$

where ρ is the correlation coefficient between the spot rate of return and future rate of return, σ_f and σ_s are the standard deviation of the spot and future rate of return respectively, and R_f is the future rate of return. It is clear that, if the futures price follows martingale process, the zero-VaR hedge ratio would be the same as the MV hedge ratio.

97. Minimum Variance Hedge Ratio

The minimum variance hedge ratio (or optimal hedge ratio) is the ratio of futures position relative to the spot position that minimizes the variance of the position. In general, this ratio is called the Johnson hedge ratio.

98. Minimum-Variance Frontier

Graph of the lowest possible portfolio variance that is attainable for a given portfolio expected return.

99. Minimum-Variance Portfolio

The portfolio of risky assets with lowest variance.

100. Mission Statement

The firm must operate to achieve a purpose or goal; otherwise, decisions will be made carelessly, allowing better informed, more serious competitors to put the firm out of business. Although many firms may have mission statement espousing goals of quality, customer service, fair prices, and so on, such qualitative statements are really only a means to an end. The firm's managers need a definite benchmark against which to evaluate the alternative means for attaining these goals.

101. Mixed Average

[See **Geometric average**]

102. Mobile Banking

Mobile banking is the act of making financial transactions on a mobile device (cell phone, tablet, etc.). This activity can be as simple as a bank sending fraud or usage activity to a client's cell phone or as complex as a client paying bills or sending money abroad. Advantages to mobile banking include the ability to bank anywhere and at any time. Disadvantages include security concerns and a limited range of capabilities when compared to banking in person or on a computer.

103. Mode

The mode is the most likely value of the distribution or the outcome with the highest probability.

104. Model Validation

Model validation is defined within regulatory guidance as the set of processes and activities intended to verify that models are performing as expected, in line with their design objectives, and business uses. It also identifies potential limitations and assumptions, and assesses their possible impact. Generally, validation activities are performed by individuals independent of model development or use. Model validation consists of four crucial elements which should be considered (i) conceptual design, (ii) system validation, (iii) data validation and quality assessment, and (iv) process validation.

105. Modern Portfolio Theory (MPT)

Principles underlying analysis and evaluation of rational portfolio choices based on risk-return trade-offs and efficient diversification. [See **Markowitz model**]

106. Modified Accelerated Cost Recovery System [MACRS]

MACRS depreciates assets by an accelerated method. In essence, it uses the double declining balance method until it becomes advantageous to use straight-line depreciation over the asset’s remaining life.

To ensure some uniformity, MACRS assigns assets to classes, as shown in Table 6. Assets in the 27.5- or 31.5-year classes must be depreciated using the straight-line method over the appropriate number of years. Additionally, with some exceptions, MACRS follows a half-year convention; the asset receives a half-year’s worth of depreciation in the year it is acquired, regardless of when it is actually purchased. Thus, assets in the 3-year class are actually depreciated over 4 years; the owner writes off a half-year of depreciation in the Year 1, a full year of depreciation in each of Years 2 and 3, and the remaining half-year of depreciation in year 4.

Annual accelerated depreciation percentages are given in Table 7. To determine an assets’ annual depreciation expense, the cost of the asset is multiplied by the percentage for the appropriate asset class and the appropriate year.

For example, for an asset in the 3-year class that originally cost \$50,000, the first year’s depreciation is $\$50,000 \times 0.3333 = \$16,665$; the second year’s depreciation

Table 6 Depreciation classes

3-year class	Designated tools and equipment used in research
5-year class	Cars, trucks, and some office equipment, such as computers and copiers
7-year class	Other office equipment and industrial machinery
10-year class	Other long-lived equipment
27.5-year class	Residential real estate
31.5-year class	Commercial and industrial real estate

Table 7 MACRS percentages

Asset class				
Year of ownership	3-year	5-year	7-year	10-year
1	33.33%	20.00%	14.29%	10.00%
2	44.45	32.00	24.49	18.00
3	14.82	19.20	17.49	14.40
4	7.40	11.52	12.49	11.52
5		11.52	8.93	9.22
6		5.76	8.93	7.37
7			8.93	6.55
8			4.45	6.55
9				6.55
10				6.55
11				3.29

is $\$50,000 \times 0.4445 = \$22,225$; the third year's depreciation is $\$50,000 \times 0.1482 = \$7,410$; and the final year's depreciation is $\$50,000 \times 0.0740 = \$3,700$.

107. Modified Duration

A modification to the standard duration measure so that it more accurately describes the relationship between proportional changes in a bond price and absolute changes in its yield. The modification takes account of the compounding frequency with which the yield is quoted. Modified duration (*MD*) is Macaulay's duration (*D*) divided by one plus the prevailing interest rate (*R*) on the underlying instrument, that is, $MD = \left(\frac{D}{1+R} \right)$.

108. Modified Exchange Standard

A system of currency exchanges and international payments in which foreign currencies were linked to gold and the US dollar, with the price of gold in terms of US dollars remaining fixed.

109. Modigliani and Miller (M&M) Proposition I

Without Taxes

M&M Proposition I (Modigliani and Miller 1958) makes several assumptions. First, it assumes that the firm pays no taxes. Second, it assumes that investors can borrow and lend money at the same interest rate as corporations. Thus, the firm and its investors can buy and sell securities at zero cost. Third, M&M Proposition I holds the firm's current and future real investment decisions constant. In other words, it assumes that all investors have the same expectations about the firms' future earnings and risk and investors can classify the firm with other corporations of similar business risk. Related to this idea is the assumption that future financing decisions will not affect the firm's investments in assets.

Under these assumptions, the value of the firm is unaffected by its specific capital structure. Through the use of personal borrowing or lending, investors can adjust their exposure to a firm's leverage to reflect their personal preferences.

The guiding force behind M&M is *arbitrage*. Should one strategy result in a higher value, investors will execute that strategy until the resulting supply and demand forces bring the values into line. For example, suppose there are two firms with identical business risk but Firm L is levered and Firm U has no debt. If Firm L were to have a higher value than Firm U, investors could use **homemade leverage** by borrowing and purchasing shares of Firm U to obtain levered cash flows more cheaply than if they purchased the shares of levered Firm L outright. [See also **Homemade leverage**] Investors purchases of the cheaper shares of Firm U would cause their price to rise until the value of Firm U was equal to that of Firm L.

Capital structure does not affect firm value under M&M Proposition I, as individual investors can costlessly adjust the effects of the corporation's leverage to suit themselves. If investors expect the firm to generate an operating profit of $E(EBIT)$ in perpetuity, the value of the levered firm (V_L) or the unlevered firm (V_U) will be that of a perpetuity:

$$V_L = V_U = \frac{E(EBIT)}{WACC} = \frac{E(EBIT)}{k_u} \quad (19)$$

where $WACC$ represents the investors' required return (and hence the firm's cost of capital) for firms with this class of business risk; k_u represents a cost of unlevered equity.

The process of homemade leverage also can work in reverse. If a firm increases its debt-to-equity ratio to a level considered too risky by an investor, the investor can reverse the capital structure to receive cash flows identical to those before the firm took on the extra debt.

Under M&M Proposition I's assumptions, the only factors that affect firm value are the firm's level of expected operating income and its business risk, which is measured by the variation in operating income. Firm value is not affected by changes in financing strategy, since whatever the firm does, investors can use homemade leverage (or unleverage) to change the cash flows they receive.

With Tax

When there is a tax, then Eq. 19 should be redefined as

$$V_L = V_U + (T)(D), \quad (20)$$

where:

- V_L = market value of levered firm
- V_U = market value of unlevered firm
- T = marginal corporate tax rate

- D = total debt
- $(T)(D)$ = tax shield value

110. Modigliani and Miller (M&M) Proposition II

Without Taxes

M&M Proposition II defines a firm's cost of equity capital (and shareholders' required rate of return) in a world of no taxes, not transaction costs, and constant real investment decisions. The cost of equity capital for an unlevered firm (k_u), or a firm with no debt in its capital structure, is simply its expected level of operating income ($EBIT$) divided by its assets (TA):

$$k_u = E(EBIT)/TA = E(EBIT)/Equity$$

M&M show that the cost of equity for a levered firm (k_e) is:

$$k_e = k_u + (k_u - k_d)(D/E), \quad (21)$$

where k_d is the cost of debt; $(k_u - k_d)(D/E)$ is the financial risk premium; and D/E is the debt equity ratio. The equation says that the cost of levered equity equals the cost of unlevered equity plus a risk premium to compensate shareholders for financial risk.

With Tax

If there exist a tax, then Eq. 21 should be rewritten as

$$k_e = k_u + (k_u - k_d)(D/E)(1 - T);$$

where T represents marginal corporate tax rate.

111. Modified Internal Rate of Return

Modified internal rate of return (MIRR) rankings of mutually exclusive projects with comparable sized initial investments will agree with the NPV rankings of those projects. Additionally, the MIRR calculation always gives a single solution.

MIRR is calculated by way of a three-step solution:

1. Using the minimum required rate of return as the discount rate, find the *present* value of all cash outflows (for a conventional project, this will be just the initial cost of the project). This step converts all of the cash outflows into a lump-sum present value at Time 0.
2. Using the required return as the reinvestment or compounding rate, compute the *future* value of each cash inflow as of the end of the project's life. Add the future values together; this sum is sometimes called the terminal value. This step converts all cash inflows into a lump-sum future value at Time N .

3. Find the discount rate that allows the present value of the cash outflows to grow equal to the terminal value; this discount rate is the modified internal rate of return.

The decision rule for MIRR is similar to that for IRR: A project is acceptable if its MIRR exceeds the project's minimum required return. A drawback to the MIRR is that it, like the IRR, is a relative measure of attractiveness; it does not indicate the dollar amount by which projects change shareholder wealth.

112. Moments (of a Statistical Distribution)

Statistical distributions show the frequency at which events might occur across a range of values. The most familiar distribution is a Normal “bell shaped” curve. In general though, the shape of any distribution can be described by its (infinitely many) moments.

- (a) The **first** moment is the *mean* which indicates the central tendency.
- (b) The **second** moment is the *variance* which indicates the width.
- (c) The **third** moment is the *skewness* which indicates any asymmetric “leaning” wither left or right.
- (d) The **fourth** moment is the *kurtosis* which indicates the degree of central “peakedness” or, equivalently, the “fatness” of the outer tails.

113. Monetarist View

An approach to economy policy which contends that the money supply is a dominant influence on the price level, total spending, production, and employment in the economy.

114. Monetary Base

The sum of legal reserves in the banking system plus the amount of currency and coin held by the public.

115. Monetary Policy

Economists believe that money supply growth has broad implications for future economic growth and future levels of inflation. As a consequence, most financial managers are interested in money supply changes over time and the current status of monetary policy. This gives them information about interest rates and inflation rate trends.

Monetary policy involves the use of the **Federal Reserve Board's** powers to affect the money supply, interest rates, and aggregate economic activity. [See also **Federal Reserve Board**]

116. Monetary Policy

The Federal Reserve has four instruments of credit control with which to administer monetary policy. These four instruments are (i) open-market operations of repurchasing or selling government securities, (ii) determination of Federal Funds Rate, (iii) setting of reserve requirements for banks, and (iv) establishment of margin requirements for carrying and purchasing financial securities. These instruments generally are divided into two categories: (1) those that affect all financial markets and (2) those that affect specific areas of the economy. Instruments 1–3 are general tools, while instrument 4 is a selective tool. The principal use of these instruments is to regulate the money supply defined as M1, M2, M3, and M4. M1 is the traditional definition of the money supply: currency and coin plus demand deposits (checking accounts) at commercial banks, negotiable order of withdrawal (NOW) accounts, share draft accounts (credit unions), and savings accounts that automatically transfer funds into a checking account when the checking account is overdrawn. The two components have elements of the traditional savings account but also have readily spendable characteristics.

117. Money

A financial asset that serves as a medium of exchange and standard of value for purchases of goods and services.

118. Money Creation

The ability of banks and other depository institutions to create a deposit, such as a checking account, that can be used as a medium of exchange (to make payments for purchases of goods and services).

119. Money Market

In the money market, securities are issued or traded that mature, or come due, in 1 year or less. Examples of money market securities include US Treasury bills, corporate commercial paper, and negotiable certificates of deposit.

120. Money Market Account

An investment that is initially equal to \$1 and, at time t , increases at the very short-term risk-free interest rate prevailing at that time.

121. Money Market Deposit Accounts (MMDAs)

Deposits whose interest yields vary with market conditions and are subject to withdrawal by check.

122. Money Market Liquidity Premium

A liquidity premium is additional value demanded by investors when any given security cannot be easily and efficiently sold or otherwise converted into cash for its fair market value. When the liquidity premium is high, the asset is said to be illiquid. Investors of illiquid assets require compensation for the added risk of investing their funds in assets that may not be able to be sold for an extended period, especially since valuations can fluctuate with market effects in the interim.

123. Money Market Mutual Fund

Mutual fund that accepts customer funds and purchase short-term marketable securities.

124. Money Market Securities

The classification of a financial instrument as a marketable security typically is based upon maturity and, to a lesser extent, liquidity. Investment instruments with maturities of 1 year or less that are traded to some extent in secondary markets are called money market securities.

125. Money Multiplier

The ratio of the size of the nation's money supply to the total reserve base available to depository institutions.

126. Money Purchase Plan

A defined benefit contribution plan in which the participant contributes some part and the firm contributes at the same or a different rate. Also called an individual account plan.

127. Money Spread

[See **Spread (options)**]

128. Money Supply

The federal government's designation of certain liquid assets as money; M1A equals currency outside banks plus demand deposits; M1B equals M1A plus other checkable deposits; M2 equals M1B plus overnight RPs, savings and small time deposits, and money market funds; M3 equals M2 plus large time deposits and term RPs; L equals M3 plus other liquid assets (where RPs and repurchase agreements).

129. Money Supply Expectations Effect

A method for forecasting interest rates that compares actual growth of the nation's money supply with the market's expectation for money supply growth; for example, if expected money supply growth exceeds actual growth, interest rates will tend to fall.

130. Money Supply Income Effect

Increases and decreases in the nation's income and spending resulting in changes in the demand for money and leading to corresponding increase or decrease in the interest rates

131. Money Supply Liquidity Effect

Increases or decreases in the nation's money supply causing interest rates to move in the opposite direction (assuming money demand is unchanged).

132. Monotonicity

[See **Rank order**]

133. Monte Carlo Valuation (Simulation)

A procedure for pricing derivative claims by discounting expected payoffs, where the expected payoff is computed using simulated prices for the underlying asset.

134. Moody's Bond Rating

Aaa – bonds of highest quality

Aa – bonds of high quality

A – bonds whose security of principal and interest is considered adequate but may be impaired in the future

Baa – bonds of medium grade that are neither highly protected nor poorly secured

Ba – bonds of speculative quality whose future cannot be considered well assured

B – bonds that lack characteristics of a desirable investment

Caa – bonds in poor standing that may be defaulted

Ca – speculative bonds that are often in default

C – bonds with little probability of any investment value (lowest rating)

135. Moral Hazard

Moral hazard is the risk that a party has not entered into a contract in good faith or has provided misleading information about its assets, liabilities, or credit capacity. In addition, moral hazard also may mean a party has an incentive to take unusual risks

in a desperate attempt to earn a profit before the contract settles. Moral hazards can be present at any time two parties come into agreement with one another. Each party in a contract may have the opportunity to gain from acting contrary to the principles laid out by the agreement.

136. Moral Suasion

A monetary policy tool of the central bank in which its officers and staff try to persuade bankers and the public through speeches, conferences, and written communications to conform more closely to the central bank's goals.

137. Mortality Tables

Tables of probability that individuals of various ages will die within a year; created and used by life insurance companies.

138. Mortgage

A contract whereby a borrower provides a lender with a lien on real property as security against a loan.

139. Mortgage Banking

The business of packaging mortgage loans for sale to investors and retaining the servicing rights to the mortgages.

140. Mortgage Banks

Financial service firms that work with property developers to provide real estate financing and then place the long-term loans with long-term lenders such as insurance companies and savings banks.

141. Mortgage-Backed Security

An ownership claim in a pool of mortgages or an obligation that is secured by such a pool. These claims represent securitization of mortgage loans. Mortgage lenders originate loans and then sell packages of these loans in the secondary market. Also called a *pass-through*, because payments are passed along from the mortgage originator to the purchaser of the mortgage-backed security.

142. Mortgage Bonds

Mortgage bonds pledge real property or specific assets as security. [See **Collateral**]

143. Mortgage Securities

A debt obligation secured by a mortgage on the real property of the borrower.

144. Mortgage Servicing

The process of collecting monthly payments on mortgages, keeping records, paying the associated insurance and taxes, and making monthly payments to holders of the underlying mortgages or mortgage-backed securities.

145. Move Persistence

The statistical tendency in a time series to move on the next step in the same direction as the previous step. [See **Autocorrelation**]

146. Moving-Average

Moving-average (or rate-of-change) technicians focus on prices and/or moving averages of prices. The moving average is used to provide a smoothed stable reference point against which the daily fluctuations can be gauged. When the daily prices penetrate above the moving-average line, technicians interpret this penetration as a bearish signal. When the daily prices move downward through the moving average, they frequently fail to rise again for many months.

Moving-average analysts recommend buying a stock when: (1) the 200-day moving average flattens out and the stock's price rises through the moving average, (2) the price of a stock falls below a moving-average line that is rising, and (3) the price of a stock that is above the moving-average line falls but turns around and begins to rise again before it ever reaches the moving-average line.

Moving-average chartists recommend selling a stock when: (1) the moving-average line flattens out and the stock's price drops downward through the moving-average line, (2) a stock's price rises above a moving-average line that is declining, and (3) a stock's price falls downward through the moving-average line and turns around to rise but then falls again before getting above the moving-average line.

147. Multi-Period Capital Budgeting

Capital investments are long-term investments in which the assets involved have useful lives of multiple years. Capital budgeting is a method of estimating the financial viability of a capital investment over the life of the investment. Capital budgeting involves identifying the cash in flows and cash out flows rather than accounting revenues and expenses flowing from the investment. The cash flow expenditures associated with the actual purchase and/or financing of a capital asset are included in the analysis. Capital investments create cash flows that are often spread over several years into the future. To accurately assess the value of a capital

investment, the timing of the future cash flows are taken into account and converted to the current time period (present value).

148. Multibank Holding Company

A bank holding company that owns controlling interest in at least two commercial banks.

149. Multifactor CAPM

Generalization of the basic CAPM that accounts for extra-market hedging demands.

150. Multinational Bank

A commercial bank engaged in selling services or conducting operations in more than one country.

151. Multinational Corporation

A large company with manufacturing, trading, or service operations in several different countries.

152. Multiple Rates of Return

More than one rate of return from the same project that make the net present value of the project equal to zero. This situation arises when the IRR method is used for a project in which negative cash flows follow positive ones.

153. Multiples

Another name for price/earnings ratios.

154. Municipals

Securities issued by states, local governments, and their political subdivisions.

155. Municipal Bonds

Tax-exempt bonds issued by state and local governments, generally to finance capital improvement projects. General obligation bonds are backed by the general taxing power of the issuer. Revenue bonds are backed by the proceeds from the project or agency they are issued to finance.

156. Mutual

Depository Institutions owned by their depositors, such as savings banks and most saving and loan associations.

157. Mutual Fund

A pool of funds that is managed by an investment company. Investors in a mutual fund own shares in the fund, and the fund uses the proceeds to buy different assets. Some of the important mutual funds are money market funds, fixed –income funds, balance funds, income funds, asset allocation funds, index funds and growth funds.

158. Mutual Fund Theorem

A result associated with the CAPM, asserting that investors will choose to invest their entire risky portfolio in a market-index mutual fund.

159. Mutual Savings Bank

Firms without capital stock that accept deposits and make loans.

160. Mutually Exclusive Investment Decisions

Investment decisions in which the acceptance of a project precludes the acceptance of one or more alternative projects.

161. Mutually Exclusive Projects

Investment projects are mutually exclusive, or competing, projects when they represent different alternatives to meet the same perceived need. Since all of the alternatives seek to meet the same need, the firm will choose only the one that creates the most value for shareholders.

1.14 N

1. Naked Options

The writing of a call or put option without owning the underlying asset is known as naked options. Naked options are much riskier than the covered options. [See **Covered call**]

2. Naked Option Writing

The owner of shares of common stock can write, or create, an option and sell it in the options market, in an attempt to increase the return or income on a stock

investment. A more venturesome investor may create an option in this fashion without owning any of the underlying stock. This naked option writing exposes the speculator to unlimited risk because he or she may have to buy shares at some point to satisfy the contract at whatever price is reached. This is a serious risk if the value of the underlying asset has a high degree of variability.

3. Naked Position

A short position in a call option that is not combined with a long position in the underlying asset. An alternative to a naked position, a financial institution can adopt a covered position which is a short position combining a long position in the underlying asset.

4. Naked Writing

Selling option without an offsetting position in the underlying asset.

5. NAIC

National Association of Insurance Commission, which is an organization with no legal power but with substantial political clout. Commissioners of insurance in each state wield considerable power individually and exert influence collectively through NAIC.

6. Naïve Hedge Ratio

Naïve hedge ratio implies that \$1 of spot is hedged by \$1 of futures. Therefore, the naïve hedge ratio is equal to 1. It is well known that the optimum hedge ratio can be either larger than 1 or smaller than 1. Hence, the naïve hedge ratio is not an optimum hedge ratio.

7. Nasdaq

It represents the National Association of Securities Dealers Automated Quotation (Nasdaq). This automated quotation system is designed for the OTC market, showing current bid-asked prices for thousands of stocks.

8. Nasdaq Index

This index includes 4,000 over-the-counter (OTC) firms traded on Nasdaq market.

9. National Banks

US banking institutions that received their charter of incorporation from the Comptroller of the Currency, an agency of the US government.

10. National Credit Union Administration (NCUA)

Federal regulatory agency that oversees the activities of federally chartered credit unions.

11. National Income Accounts

A system of social accounts compiled and released quarterly by the US Department of Commerce that presents data on the nation's production of goods and services, income flows, spending, and saving.

12. Negative Covenant

Part of the indenture or loan agreement that limits or prohibits actions that the company may take.

13. Negative Pledge Clause

A negative pledge clause in a debenture agreement states that any future debt-financed asset purchases also are considered to be security for the bond, even if the assets are financed with first mortgage bonds.

14. Neglected-Firm Effect

Small firms tend to be neglected by large institutional traders. It has been found by Arbel (1985) that investment in stock of this kind in less well-known firms has generated abnormal returns. [See **January effect**]

15. Negotiable Certificates of Deposit

Negotiable certificates of deposit (CDs) are financial instruments offered by banks to customers who deposit funds for fixed periods at fixed rates of interest. CDs are issued in denominations of \$100,000 or more, with maturities ranging to several years.

Yields on CDs are higher than yields on T-bills for two reasons. First, CDs are substantially less liquid than T-bills (their secondary market is very thin). Second, CDs have higher default risk because they represent unsecured debt obligations of the issuing banks. However, the spread between CD and T-bill yields varies depending upon economic conditions, supply and demand forces, and investor attitudes.

16. Negotiable Order of Withdrawal (NOW)

In 1980, Congress enacted depository institutions deregulation and monetary control act (DIDMCA). Titled III of DIDMCA authorized banks and financial

institution offered interest-bearing transactions account. In banks and thrifts, they call this kind of account NOW account.

17. Negotiated Credit

Short-term bank credit is particularly important to the smaller company. Many large, well-established companies make little use of bank credit. When they need working capital above what is available as trade credit – that is, when they need negotiated credit, the term given to all credit that arises from a formal negotiation of funds – they can get attractive terms by borrowing directly from the capital market. This borrowing usually takes the form of selling commercial paper.

18. Negotiated Markets

Institutional mechanisms set up by society to make loans and trade securities where the terms of trade are set by direct bargaining between a lender and a borrower.

19. Negotiated Offer

The issuing firm negotiates a deal with one underwriter to offer a new issue rather than taking competitive bidding.

20. Negotiation

One technique used in business combinations is direct negotiation between the management teams and the boards of directors of the two firms. After negotiations have been worked out, the plans are presented to both shareholder groups for approval.

Negotiation must identify what the firms will exchange, at what prices, and the method of payment. Assume that Firms A and B negotiate so that Firm B acquires all the assets (except cash) of Firm A and pays for these assets with its own cash. Now Firm A has cash as its only asset, and it may pay off its creditors and distribute any remaining cash as a liquidating dividend to its shareholders. If, however, Firm B pays for the assets of Firm A with its own shares of stock, then Firm A may sell off the stock and distribute the cash or distribute the stock directly to its shareholders. Note that the effect of these negotiations on the balance sheet of Firm B is an increase in the assets account, to reflect the acquired assets, and a decrease in cash or an increase in the capital accounts, to reflect the method of payment.

Assume now that Firm B acquires the common stock of Firm A (and not the assets directly). Firm B may acquire the shares for cash, either in exchange for some of its shares or by some more complex plan. In the extreme case in which the shareholders of Firm A surrender all their shares for shares of Firm B, Firm A ceases to exist and Firm B assumes all the assets and liabilities of Firm A. State laws specify that once a certain percentage of A's shareholders agree to an exchange of shares, all

shareholders must comply. Holdout shareholders of Firm A may go to the courts to earn a fair price for their shares in the even that they are not satisfied with the negotiated price. In a less extreme case, Firm B may acquire less than all of the shares of Firm A and maintain an interest in Firm A. In this case, the shares of Firm A appear as an investment on the balance sheet of Firm B.

21. Net Cash Balance

Beginning cash balance plus cash receipts minus cash disbursements.

22. Net Float

Sum of disbursement float and collection float. [See **Float**]

23. Net Interest Margin

Ratio of net interest income to total earning assets; used to evaluate profitability of banks.

24. Net Investment

Gross, or total, investment minus depreciation.

25. Net Operating Losses (NOL)

Losses that a firm can take advantage of to reduce taxes.

26. Net Overhead Burden

Difference between noninterest expense and noninterest income as a fraction of total bank assets.

27. Net Payoff

Another term for profit.

28. Net Present Value

The net benefit, or the net present value (NPV), of an investment is the present value of a project's cash flows minus its cost. The present value of the expected cash flows from a project is found by discounting each cash flow to the present. The *net present value* (NPV) is defined as:

Net present value = Present value of the expected cash flows – Cost the of project

More formally, NPV is

$$\begin{aligned}\text{NPV} &= \frac{\text{CF}_1}{(1+r)^1} + \frac{\text{CF}_2}{(1+r)^2} + \dots + \frac{\text{CF}_N}{(1+r)^N} - I \\ &= \text{CF}_1[\text{PVIF}(r, 1)] + \text{CF}_2[\text{PVIF}(r, 2)] + \dots + \text{CF}_N[\text{PVIF}(r, N)],\end{aligned}$$

where:

- CF_t = Annual cash flow generated by the project in period t ($t = 1, 2, \dots, N$)
- $\text{PVIF}(r, t)$ = Present value factor for r percent in period t
- I = Initial cost of the project
- N = Expected life of the project
- r = Required rate of return used to discount the cash flows

It is a “net” present value in that it subtracts the project’s investment cost from the present value of the project’s expected cash flows.

29. Net Present Value Profile

Management may want to assess the sensitivity of a project’s NPV to the required rate of return. An NPV profile shows this relationship in a graph of project NPVs for different values of the discount rate. The calculations and graphing of an NPV profile can be handled easily by a spreadsheet program.

If the resulting NPV profile shows a steeply sloped curve, then the NPV of the project under consideration is sensitive to the discount rate assumption. In such a case, management should carefully assess the project’s required return. If the NPV profile is sloped gradually, then the project’s impact on shareholder wealth is not very sensitive to changes in the discount rate.

30. Net Present Value Rule

An investment is worth making if it has a positive NPV. If an investment’s NPV is negative, it should be rejected.

31. Net Working Capital

Net working capital, the difference between current assets and current liabilities, is a financial indicator that can be used in conjunction with ratio analysis to gauge a firm’s liquidity. An increase in net working capital is a net investment in the firm’s current assets; and an increase in an asset is considered a use of cash. A decrease in net working capital is a divestment of assets, that is, a source of cash. In general, an abundance of net working capital suggests that the firm has ample liquidity to meet its short-term obligations.

$$\text{Net working capital} = \text{Current assets} - \text{Current liabilities}$$

But this may not always be the case. In fact, one of the objectives of short-term financial planning is to reduce excess or redundant working capital to a minimum, since carrying these idle assets has both an explicit and implicit cost.

32. Netting

The practice of offsetting promised interest payments with promised interest receipts and transferring the difference with an interest rate swap. [See **Interest rate swap**] There are at least three types of netting:

- (a) *Close-out netting*: In the event of counterparty bankruptcy, all transactions or all of a given type are netted at market value. The alternative would allow the liquidator to choose which contracts to enforce and which to not to (and thus potentially “cherry pick”) There are international jurisdictions where the enforceability of netting in bankruptcy has not been legally tested.
- (b) *Netting by novation*: The legal obligation of the parties to make required payments under one or more series of related transactions is canceled and a new obligation to make only the net payment is created.
- (c) *Settlement or payment netting*: For cash settled trades, this can be applied either bilaterally or multilaterally and on related or unrelated transactions.

33. Net Worth

Owner's (stockholders') equity in a firm.

34. Newton–Raphson Method

The Newton–Raphson procedure is designed to solve an equation of the form $f(x) = 0$. It starts with a guess of the solution: $x = x_0$. It then produces successively better estimates of the solution: $x = x_1, x = x_2, x = x_3, \dots$ using the formula $x_{i+1} = x_i - f(x_i)/f'(x_i)$. Usually, x_2 is extremely close to the true solution.

35. No-Arbitrage Assumption

The assumption that there are no arbitrage opportunities in market prices.

36. No-Arbitrage Interest Rate Model

A model for the behavior of interest rates that is exactly consistent with the initial term structure of interest rates. [See **Term structure of interest rates**]

37. No Loan Fund

A mutual fund that does not charge a regular sales commission or sale charge. In other words, there are no front end sales charges.

38. Nominal Cash Flow

A cash flow expressed in nominal terms if the actual dollars to be received (or paid out) are given.

39. Nominal Interest Rate

When the bond interest rate is quoted as an APR, it is called a nominal interest rate or **stated annual interest rate**. Given an annual percentage rate, the periodic interest rate is APR/m , where m represents the number of periods or cash flows in a year. Since APR assumes no period-by-period compounding of cash flows, it fails to account for interest-on-interest.

40. Nominal Risk-Free Interest Rate

Potential savers have little incentive to invest unless their expected returns include some protection against expected inflation. To try to protect themselves from a loss of purchasing power, investors will demand a return that reflects inflationary expectations. This return is called the nominal risk-free interest rate; it represents the observed or published return on a risk-free asset

The nominal risk-free rate depends upon: the **real risk-free rate** and the expected inflation rate. [See also **Real risk-free rate**]

$$\text{Nominal risk-free interest rate} = (1 + \text{Real risk-free interest rate}) \times (1 + \text{Expected inflation rate}) - 1$$

This equation, known as the **Fisher effect**, illustrates how the inflation rate determines the relationship between real and nominal interest rates.

Many financial analysts use the interest rate on a 1-year Treasury bill to approximate the nominal risk-free rate. The Treasury bill (T-bill) has a short time horizon and the backing of the US government, which give it an aura of safety. The 1-year T-bill rate is used because investment returns usually are stated as annual returns.

41. Nonbank Bank

A firm that either makes commercial loans or accepts deposits but does not do both. Thus, it avoids regulation as a commercial bank. In other words, it undertakes many of the activities of a commercial bank without meeting the legal definition of a bank.

42. Nonbank Subsidiary

A subsidiary of a bank holding company that is engaged in activities closely related to banking, such as leasing, data processing, factoring, and insurance underwriting.

43. Nonborrowed Reserves

The largest component of the total legal reserve of depository institutions, consisting of all those legal reserves owned by depository institutions themselves and not borrowed from the Federal Reserve banks.

44. Noncash Item

Expense against revenue that does not directly affect cash flow, such as depreciation and deferred taxes.

45. Nondebt Tax Shields

If firms pay taxes and interest is tax-deductible, firm value rises as the use of debt financing rises. But this analysis implies that there are limits to the benefits of tax-deductible debt. For example, business risk leads to variations in EBIT over time, which can lead to uncertainty about the firm's ability to fully use future interest deductions. If a firm has a negative or zero operating income, an interest deduction provides little help; it just makes the pretax losses larger. The advantage of tax-deductible interest also is reduced if the firm has tax-loss carry forwards that reduce current and future years' taxable incomes. Also, firms in lower tax brackets have less tax incentive to borrow than those in higher tax brackets.

The present value of future interest tax shields becomes even more uncertain if EBIT is affected by nondebt tax shields. In practice, firms' EBITs are reduced by various expenses, such as depreciation, depletion allowances, amortization, pension contributions, employee and retiree health-care costs, R&D, and advertising expenses. Foreign tax credits, granted by the US government to firms that pay taxes to foreign governments, also diminish the impact of the interest deduction. Thus, the tax deductibility of debt becomes less important to firms with large nondebt tax shields.

46. Nondeposit Funds

Borrowing of funds by banks (usually in the short-term money market) to supplement the funds they receive from selling deposits to the public.

47. Nondiversifiable Risk

Risk that remains after a large number of assets are combined in a portfolio. [See **Systematic risk**]

48. Noninstallment Credit

A loan that is normally paid off in a lump sum rather than in a series of installment payments.

49. Nonmarketed Claims

Claims that cannot be easily bought and sold in the financial markets, such as those of the government and litigants in lawsuits.

50. Nonnotification Financing

[See **Pledging**]

51. Nonperforming Loan

Loan for which an obligated interest payment is 90 days past due. They are placed on accrual status. Banks have traditionally stopped accruing interest when debt payments were more than 90 days past due.

52. Nonrated Bond

A bond that is not rate by Moody's, S&P, or other rating agency.

53. Nonrate Gap

Noninterest-bearing liabilities plus equity minus non-earning assets as a ratio of earning assets.

54. Nonrecourse

Holder of an obligation has no legal right to force payment on a claim.

55. Nonrecombining Tree

A binomial tree describing asset price moves in which an up move followed by a down move yields a different price than a down move followed by an up move.

56. Nonresidential Mortgages

Loans secured by business and farm properties.

57. Nonstationary Model

A model where the volatility parameters are a function of time.

58. Nonstandard Option

[See **Exotic option**]

59. Nonsystematic Risk

Nonmarket or firm-specific risk factors that can be eliminated by diversification. Also called unique risk or diversifiable risk. Systematic risk refers to risk factors common to the entire economy.

60. Normal Backwardation Theory

Normal backwardation is one of the three traditional theories used to explain the relationship between the futures price and the expected value of the spot price of the commodity at some future date. Normal backwardation suggests that the futures price will be bid down to a level below the expected spot price, and will rise over the life of the contract until maturity date. On the maturity date, futures price is equal to spot price. [See also **Expectations hypothesis** and **Contango**]

61. Normal Distribution

System metric bell-shaped frequency distribution that can be defined by its mean and standard deviation. It's a systematic distribution and therefore the skewness of normal distribution is zero. It is a continuous probability distribution that assigns positive probability to all values from $-\infty$ to $+\infty$. Sometimes called the "bell curve."

The probability density function of a normal random variable can be defined as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

where $\pi = 3.14159$, $e = 2.71828$, and $\mu(-\infty < \mu < \infty)$ and $\sigma^2(0 < \sigma^2 < \infty)$ are the mean and variance of the normal random variable X . To graph the normal curve, we must know the numerical values of π and σ^2 . [See also **Central limit theorem**]

62. Normal Market

A market where futures prices increase with maturity.

63. Note

Unsecured debt, usually with maturity of less than 15 years. Note payable is one of the liability items in the balance sheet.

64. Note Issuance Facility

An arrangement in which borrowers can issue short-term securities in their own names.

65. Notional Amount

The dollar amount used as a scale factor in calculating payments for a forward contract, futures contract, or swap.

66. Notional Principal

The principal used to calculate payments in an interest rate swap. The principal is “notional” because it is neither paid nor received.

67. Notional Value

The face value of interest rate swap contracts; a mere reference value to compute obligated interest payments.

68. NOW Account

An interest-bearing checking account available to individuals and nonprofit institutions from banks, savings and loans, and other depository institutions

69. NPV

$$\text{Net Present Value} = \text{present value of expected cash flow} \\ - \text{cost of the project}$$

[See **Net present value**]

70. NPVGO Model

A model valuing the firm in which net present value of new investment opportunities is explicitly examined. NPVGO stands for net present value of growth opportunities. This model divided the dividend growth model into two parts as value of share when firm acts as cash low plus NPV of growth opportunity. [See **Cash cow for value of a share**]

71. NSF

Not sufficient funds.

72. Numeraire

Defines the units in which security prices are measured. For example, if the price of IBM is the numeraire, all security prices are measured relative to IBM. If IBM is \$80 and a particular security price is \$50, the security price is 0.625 when IBM is the numeraire.

73. Numerical Procedure

A method of valuing an option when no formula is available.

1.15 O

1. Obligor

A party who is in debt to another. It can be either (a) a loan borrower, (b) a bond issuer, (c) a trader who has not yet settled, (d) a trade partner with accounts payable, or (e) a contractor with unfinished performance, etc. [See **Counterparty**]

2. OCC

Options Clearing Corporation. [See **Clearinghouse**]

3. Odd Lot

Stock trading unit of less than 100 shares.

4. Odd-Lot Theory

The odd-lot theory is one of several theories of contrary opinion. In essence, the theory assumes that the common mean is usually wrong and it is therefore advantageous to pursue strategies opposite to his thinking. In order to find out what the common man is doing, statistics on odd-lot trading are gathered. Most odd-lot purchases are made by amateur investors with limited resources – that is, by the common man, who is small, unsophisticated investors.

5. Off-Balance Sheet Activities

Commitments, such as loan guarantees, that do not appear on a bank's balance sheet but represent actual contractual obligations. For example, the issuance of standby letter of credit guarantee is also an off-balance-sheet activity.

6. Off-Balance-Sheet Risk

The risk incurred by a financial institution due to activities related to contingent assets and liabilities.

7. Off Balance Sheet Financing

Financing that is not shown as a liability on a company's balance sheet. In leasing, lessees needed only to report information on leasing activities in the footnotes of their financial statements. Thus, leasing led to off-balance-sheet financing.

8. Off-Market Swap

Swaps that have non-standard terms that require one party to compensate another. Relaxing a standardized swap can include special interest rate terms and indexes as well as allowing for varying notional values underlying the swap.

9. Offer Price

The price that a dealer is offering to sell an asset. It is an ask price.

10. Official Reserve Transactions

Transfer of the ownership of gold, convertible foreign currencies, deposits in the International Monetary Fund, and Special Drawing Rights (SDRs) used by a nation to settle a deficit in its balance of payments.

11. One Bank Holding Company

A holding company that owns or controls only one commercial bank.

12. One-Factor APT

A special case of the arbitrage pricing theory (APT) that is derived from the one-factor model by using diversification and arbitrage. It shows the expected return on any risky asset is a linear function of a single factor. The CAPM can be expressed as one-factor APT in which a single factor is the market portfolio.

13. On-the-Run Issue

The most recently issues US Treasury security. It is considered to be the actively traded issue.

14. Open (Good-Till-Canceled) Order

A buy or sell order remaining in force for up to 6 months unless canceled.

15. Open Account

A credit account for which the only formal instrument of credit is the invoice.

16. Open Contracts

Contracts that have been bought or sold without the transactions having been completed by subsequent sale or purchase, or by making or taking actual delivery of the financial instrument or physical commodity. Measured by “open interest,” as reported in the press.

17. Open Interest

The quantity of a derivatives contract that is outstanding at a point in time. (One long and one short position count as one unit outstanding.)

18. Open Market

Institutional mechanism created by society to make loans and trade securities in which any individual or institution can participate.

19. Open Market Operation

Open market operations are the Fed’s most frequently used monetary policy tool. The Fed buys and sells securities (usually Treasury bills) with other market participants. When it purchases government securities in the open market, the Fed trades dollars for securities. The seller deposits these dollars in a bank, thereby increasing the bank’s reserves from which it can make loans. Through a multiplier process, the open market purchase boosts deposits in the US banking system and the money supply rises. An open market sale of securities by the Fed has the opposite effect, reducing the level of loanable funds in the banking system, and therefore the money supply.

20. Open Market Repurchase

A firm can reacquire its stock through an open market repurchase. Acting through a broker, the corporation purchases shares in the secondary market just like any other investor. A corporation usually announces its intention to engage in an open market repurchase in advance, although the exact amount of shares repurchased and the actual days of the transactions are not known.

21. Open Outcry

A system of trading in which buyers and sellers in one physical location convey offers to buy and sell by gesturing and shouting.

22. Open-End (Mutual) Fund

A fund that issues or redeems its own shares at their net asset value (NAV). This kind of fund provides opportunities for small investors to invest in financial securities and diversify risk.

23. Operating Activities

Sequence of events and decisions that create the firm's cash inflows and cash outflows. These activities include buying and paying for raw materials, manufacturing and selling a product, and collecting cash.

24. Operating Cash Flow

Earnings before interest and depreciation minus taxes. It measures the cash generated from operations not counting capital spending or working capital requirements.

25. Operating Cycle

When a firm is functioning efficiently, its operating cycle moves through four stages: (1) converting cash to inventory, (2) converting inventory to sales, (3) converting sales to accounts receivable, and (4) converting accounts receivable to cash. This operating cycle can be very simple or quite complex. A cash flow timeline can depict the most complex as well as the simplest situation.

Two financial ratios, the receivable collection period and the inventory conversion period, help the manager to quantify the operating cycle. The average days of accounts receivable is defined as accounts receivable divided by sales per day:

$$\text{Receivable collection period} = \frac{\text{Accounts receivable}}{\text{Sales}/365 \text{ days}}$$

The inventory conversion period is defined as inventory divided by cost of goods sold per day:

$$\text{Inventory conversion period} = \frac{\text{Inventory}}{\text{Cost of goods sold}/365 \text{ days}}$$

Adding these two ratios together gives us the length of a firm's operating cycle:

Operating cycle = Receivables collection period + Inventory conversion period

The operating cycle measures conversion of current assets to cash. [See also **Cash conversion cycle**]

26. Operating Income

Sum of interest income and non-interest income for a financial institution. For a nonfinancial institution, it represents the net sale minus cost of goods sold.

27. Operating Lease

An operating lease is a shorter-term lease than for instance, a leveraged lease or a sale and lease-back agreement, which may be cancelled at the lessee's option. An operating lease does not satisfy any of the four financial lease criteria. [See **Capital lease**] The lessor typically must maintain and service the asset. Computers, photocopiers, and trucks often are acquired under the terms of an operating lease.

28. Operating Leverage

A firm's business risk is affected by its level of fixed costs, or in financial terminology, its operating leverage. Operating leverage magnifies the effect of changing sales to produce a percentage change in EBIT larger than the change in sales, assuming constant profit margins. It is a business risk measure. [See **Degree of operating leverage**]

29. Operating Profitability

Operating profit is the profit a firm receives from its normal business activities. Derived from gross profit, operating profit reflects the residual income that remains after accounting for all the costs of doing business. In addition to COGS, this includes fixed-cost expenses such as rent and insurance, variable expenses, such as shipping and freight, payroll and utilities, as well as amortization and depreciation of assets. All the expenses that are necessary to keep the business running must be included.

30. Operational Risk

Operational risk summarizes the uncertainties and hazards a company faces when it attempts to do its day-to-day business activities within a given field or industry. A type of business risk, it can result from breakdowns in internal procedures, people and systems – as opposed to problems incurred from external forces, such as political or economic events, or inherent to the entire market or market segment, known as systematic risk. Operational risk can also be classified as a variety of unsystematic risk, which is unique to a specific company or industry.

31. Opportunity Cost

From economics, we know that an opportunity cost is the cost of passing up the next best alternative. For example, the opportunity cost of a building is its market value. By deciding to continue to own it, the firm is foregoing the cash it could receive from selling it. Economics teaches the TINSTAAFL principle: “There is no such thing as a free lunch.” Capital budgeting analysis frequently applies this principle to existing assets.

If a firm is thinking about placing a new manufacturing plant in a building it already owns, the firm cannot assume that the building is free and assign it to the project at zero cost. The project’s cash flow estimates should include the market value of the building as a cost of investing since this represents cash flows the firm will not receive from selling the building.

32. Opportunity Set

The possible expected return – standard deviation pairs of all portfolios that can be constructed from a set of assets. Also called a feasible set.

33. Optimal Capital Structure

Capital structure is measured by total debt divided by total assets. If a company increases their debt financing, it creates a tax shield advantage. However, the financial risk can increase. If the trade-off between the tax shield advantage and financial risk reaches optimum conditions, then the company has an optimal capital structure.

34. Optimal Cash Balance

Based upon Baumol’s economic order quantity (EOQ) model the total cost of cash balances can be defined as:

$$\begin{aligned}\text{Total costs} &= \text{Holding costs} + \text{Transaction costs} \\ &= \frac{C}{2}r + \frac{T}{C}F,\end{aligned}$$

where:

C = Amount of cash raised by selling marketable securities or borrowing

$\frac{C}{2}$ = Average cash balance

r = Opportunity cost of holding cash (the foregone rate of return on marketable securities)

T = Total amount of new cash needed for transaction over entire period (usually 1 year)

$\frac{T}{C}$ = Number of transactions

F = Fixed cost of making a securities trade or borrowing money

The minimum total costs are obtained when C is set equal to C^* , the optimal cash balance. C^* is defined as follows:

$$C^* = \sqrt{\frac{2FT}{r}}$$

C^* = Optimal amount of cash to be raised by selling marketable securities or by borrowing.

[See also **Baumal's economic order quantity model**]

35. Optimal Dynamic Capital Budgeting Decision

Under certain conditions, a budgeting mechanism with threshold division of financing is optimal. Headquarters: (i) allocate a spending account to the manager and accumulate it over time; (ii) set a threshold, such that projects below it are financed from the account, while projects above are audited and financed fully by headquarters. Extensions generate new implications for the design of capital allocation processes.

36. Optimal Financial Policy

To develop a financial policy, a company should analyze its future cash requirements to set an appropriate level of indebtedness. Cash needs should be linked to capital requirement and acquisition plans and should be determined in the context of the company's business environment. As a consequence, an adequate level of financial flexibility will be necessary to match the projected cash flows and corporate strategy of the firm. In general, the optimal financial policy emphasizes the existence of the optimal capital structure. If a company increases their debt, the financial risk of the firm will increase. However, debt can create a tax shield for a company. The trade-off between financial risk and the tax shield benefit will affect the determination of the optimal capital structure.

37. Optimal Risky Portfolio

An investor's best combination of risky assets to be mixed with safe assets to form the complete portfolio.

38. Optimum Mean-Generalized Semivariance Hedge Ratio

Chen et al. (2001) extend the GSV hedge ratio to a Mean-GSV (M-GSV) hedge ratio by incorporating the mean return in the derivation of the optimal hedge ratio.

The M-GSV hedge ratio is obtained by maximizing the following mean-risk utility function, which is similar to the conventional mean-variance based utility function:

$$U(R_h) = E[R_h] - V_{\delta, \alpha}(R_h)$$

where R_h is the rate of return for a hedged portfolio. This approach to the hedge ratio does not use the risk aversion parameter to multiply the GSV as done in conventional mean-risk models. This is because the risk aversion parameter is already included in the definition of the GSV, $V_{\delta, \alpha}(R_h)$. As before, the M-GSV hedge ratio would be the same as the GSV hedge ratio if the futures price follows a pure martingale process.

39. Optimum Mean-MEG Hedge Ratio

Instead of minimizing the MEG coefficient, Kolb and Okunev (1993) alternatively consider maximizing the utility function defined as follows:

$$U(R_h) = E(R_h) - \Gamma_v(R_h), \quad (a)$$

The hedge ratio based on the utility function defined by Eq. (a) is denoted as the M-MEG hedge ratio. The difference between the MEG and M-MEG hedge ratios is that the MEG hedge ratio ignores the expected return on the hedged portfolio. Again, if the futures price follows a martingale process (i.e., $E(R_h) = 0$), then the MEG hedge ratio would be the same as the M-MEG hedge ratio.

40. Optimum Mean-Variance Hedge Ratio

Various studies have incorporated both risk and return in the derivation of the hedge ratio. For example, we can derive the optimal hedge ratio that maximizes the following quadratic utility function:

$$\text{Max}_{C_f} V(E(R_h), \sigma; A) = E(R_h) - 0.5\sigma_h^2 \quad (22)$$

It is clear that this quadratic utility function incorporates both risk and return. Therefore, the hedge ratio based on this utility function would be consistent with the mean-variance framework. The optimal number of futures contract and the optimal hedge ratio are respectively given by:

$$h_2 = -\frac{C_f^* F}{C_s S} = -\left[\frac{E(R_f)}{A\sigma_f^2} - \rho \frac{\sigma_s}{\sigma_f} \right] \quad (23)$$

where C_s is equal to units of a long spot position, C_f is equal to units of a short futures position, S is equal to the spot prices, F is equal to the futures prices, ρ is equal to the correlation coefficient between R_s and R_f , and σ_s and σ_f are standard deviations of R_s

and R_f , respectively. $E(R_f)$ is equal to the expected rate of return of the futures contract.

One problem associated with this type of hedge ratio is that in order to derive the optimum hedge ratio, we need to know the individual's risk aversion parameter. Furthermore, different individuals will choose different optimal hedge ratios, depending on the values of their risk aversion parameter. Since the Mean-Variance hedge ratio is easy to understand and simple to compute, it will be interesting and useful to know under what condition the above hedge ratio would be the same as the MV hedge ratio. It can be seen from Eqs. 22 and 23 that if $A \rightarrow \infty$ or $E(R_f) = 0$, then h_2 would be equal to the minimum-variance hedge ratio. The first condition is simply a restatement of the infinitely risk-averse individuals. However, the second condition does not impose any condition on the risk-averseness, and this is important. It implies that even if the individuals are not infinitely risk averse, then the mean-variance hedge ratio would be the same as the optimal mean-variance hedge ratio if the expected return on the futures contract is zero (i.e., futures prices follow a simple martingale process). Therefore, if futures prices follow a simple martingale process, then we do not need to know the risk aversion parameter of the investor to find the optimal hedge ratio.

41. Option

A right – but not an obligation – to buy or sell under-lying assets at a fixed price during a specified time period.

42. Option Bound

Boundary conditions are the maximum and minimum values used to indicate where the price of an option must lie. Boundary conditions are used to estimate what an option may be priced at, but the actual price of the option may be higher or lower than what is set as the boundary condition. For all options contracts, the minimum boundary value is always zero, since options cannot be priced at negative money. Meanwhile, maximum boundary values will differ depending on the whether the option is a call or put, and if it is an American or European style option. Option bound can be obtained by either the linear programming method or the stochastic dominance method.

43. Option Class

All options of the same type (call or put) on a particular stock.

44. Option Contracts

Agreements between contract writers and contract buyers to accept delivery of (“call”) securities or place with buyers (“put”) securities at a specified price on or before the date the contracts expire.

45. Option Elasticity

The percentage increase in an option's value given a 1% change in the value of the underlying security.

46. Option Overwriting

Selling a call option against a long position in the underlying asset.

47. Option (Maximum) Pain

Option pain is based on a controversial theory that the price of the underlying stock of options contracts can be, or is, manipulated in some way to ensure that the most amount of contracts expire out of the money. The price of the underlying stock which would cause the most amount of loss to holders of contracts is known as max pain or option pain. It is a theory that is easy to discount for many, although there are those that believe it is possible to profit from option pain.

48. Option Premium

The price of an option. [See **Option pricing equation**]

49. Option Pricing Equation

An exact formula for the price of a call option. The formula requires five variables: the risk-free interest rate, the variance of the underlying stock, the exercise price, the price of the underlying stock and the time to expiration.

$$C = SN(d_1) - Xe^{-rt}N(d_2),$$

where:

$$d_1 = \frac{\ln(S/X) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

C = Current call option value

S = Current stock price

$N(d)$ = The probability that a random draw from a standard normal distribution will be less than d ; In other words, it equals the area under the normal curve up to d

X = Exercise price

$e = 2.71828$, the base of the natural log function

r = Risk-free interest rate

$\ln$ = Natural logarithm function

σ = Standard deviation of the annualized continuously compounded rate of return of the stock

Like all models, the Black-Scholes formula is based on some important underlying assumptions:

1. The stock will pay no dividends until after the option expiration date.
2. Both the interest rate, r , and variance rate, σ^2 , of the stock are constant.
3. Stock prices are continuous, meaning that sudden extreme jumps such as those in the aftermath of an announcement of a takeover attempt are ruled out.

50. Option Series

All options of a certain class with the same strike price and expiration date.

51. Option Theoretic

An approach to estimating the expected default frequency of a particular firm. It applies Robert Merton's model-of-the-firm which states that debt can be valued as a put option of the underlying asset value of the firm. [See **KMV (1993) "Credit Monitor Overview" San Francisco: KMV Corporation**]

52. Option Writer

The party with a short position in the option.

53. Option-Adjusted Spread (OAS)

A procedure for valuing prepayment risk associated with mortgage-backed securities that recognize the magnitude and timing of prepayments and required return to an investor. This kind of model uses option pricing theory to figure the fair yield on pass-throughs and, in particular, the fair yield spread of pass-throughs over treasuries. These so-called option-adjusted spread (OAS) models focus on the prepayment risk of pass-throughs as the essential determinant of the required yield spread of pass-through bonds over treasuries.

Stripped to its basics, the option model views the fair price on a pass-through such as a GNMA (Ginnie) bond as being decomposable into two parts:

$$P_{\text{GNMA}} = P_{\text{TBOND}} - P_{\text{PREPAYMENT OPTION}}$$

That is, the value on a GNMA bond to an investor (P_{GNMA}) is equal to the value of a standard noncallable Treasury bond of the same duration (P_{TBOND}) minus the value of the mortgage holder's prepayment call option ($P_{\text{PREPAYMENT OPTION}}$). Specifically, the ability of the mortgage holder to prepay is equivalent to the bond investor writing a call option on the bond and the mortgagee owning or buying the option. If interest rates fall, the option becomes more valuable as it moves into the money and more mortgages are prepaid early by having the bond called or the prepayment option exercised. This relationship can also be thought of in the yield dimension:

$$Y_{\text{GNMA}} = Y_{\text{TBOND}} + Y_{\text{OPTION}}$$

The investors' required yield on a GNMA (Y_{GNMA}) should equal the yield on a similar duration T-bond (Y_{TBOND}) plus an additional yield for writing the valuable call option (Y_{OPTION}). That is, the fair yield spread or **option-adjusted spread (OAS)** between GNMA's and T-bond should reflect the value of this option.

54. Option-Like Securities

Option-like securities include convertible bonds, warrants, executive options, callable bonds, and risky debt. All of these securities can be evaluated by the option pricing theory.

55. Order Book Official

[See **Board broker**]

56. Order Statistics

The n draws of a random variable sorted in ascending order. It is nonparametric statistics.

57. Ordinal Utility

An ordinal utility implies that a consumer needs not be able to assign numbers that represent (in arbitrary unit) the degree or amount of utility associated with commodity or combination of commodity. The consumer can only rank and order the amount or degree of utility associated with commodity.

58. Organized Exchanges

The organized exchanges have physical locations where brokers act as agents; they help their client buy and sell securities by matching orders. The New York stock Exchange (NYSE) is the largest organized exchange in the United States.

59. Original-Issue-Discount-Bond

A bond issued with a discount from par value. Also called a deep-discount or pure discount bond.

60. Origination Fee

Fee charged by a lender for accepting the initial loan application and processing the loan.

61. Originator

The financial institution that extends credit on a facility which may later be held by another institution through, for instance, a loan sale. Originator can change origination fee. [See **Facility**]

62. Out of the Money

The owner of a put or call is not obligated to carry out the specified transaction, but has the *option* of doing so. If the transaction is carried out, it is said to have been *exercised*. If the call option is *out of the money* – that is, the stock is trading at a price below the exercise price – you certainly would not want to exercise the option, as it would be cheaper to purchase stock directly.

63. Out-of-the-Money Option

Either (a) a call option where the asset price is less than the strike price or (b) a put option where the asset price is greater than the strike price.

64. Out Performance Option

An option in which the payoff is determined by the extent to which one asset price is greater than another asset price, called the benchmark. It is also called exchange option. [See **Exchange option**]

65. Outsourcing

Buying services from third-part vendor. For example, some banks might out-source their data processing.

66. Overdraft

Depositor writing a check for an amount greater than the deposit balance.

67. Overhead

Expenses that generally do not vary with the level of output.

68. Oversubscribed Issue

Investors are not able to buy all the shares they want, so underwriters must allocate the shares among investors. This occurs when a new issue is under priced.

69. Oversubscription Privilege

Allows shareholders to purchase unsubscribed shares in a rights offering at the subscription price. This kind of privilege makes it unlikely that the corporate issuer would need to turn to its underwriter for help.

70. Over-the-Counter Market

The over-the-counter (OTC) market is a telecommunications network of dealers who provide liquidity to investors by their willingness to “make markets” in particular securities. When an investor wants to purchase a security, a dealer firm will sell it (at a price equal to the “ask” price) from its own inventory of securities; if an investor wants to sell, the dealer will purchase the security (at the “bid” price) and hold it in inventory. A source of dealer profit is the spread, or difference between the bid and ask price.

1.16 P

1. PAC

Planned amortization class such as collateralized mortgage option (CMO) – A security that is retired according to a planned amortization schedule, while payments to other classes of securities are slowed or accelerated. The objective is to ensure that PACs exhibit highly predictable maturities and cash flows.

2. Package

A derivative that is a portfolio of standard calls and puts, possibly combined with a position in forward contracts and the asset itself.

3. Pac-Man Strategy

In a pac-man strategy, the target firm tries to turn the tables and take over the hostile bidder. [See also **Tender offer**]

4. Par Bond

A bond for which the price at issue equals the maturity value.

5. Par Coupon

The coupon rate on a par bond.

6. Par Value

The face value of a bond is called the par value. Generally, this is the amount of money that the issuer has initially borrowed and promised to repay at a future maturity date. Most US corporate bonds have a par value of \$1,000 per bond.

7. Par Yield

The coupon on a bond that makes its price equal the principal.

8. Parallel Shift in the Yield Curve

A change in interest rate where rates at all maturities change by the same amount, in the same direction, at the same time. This never actually occurs.

9. Parent Company

A firm that owns controlling interest in the stock of another firm.

10. Partial Expectation

The sum (or integral) of a set of outcomes times the probability of those outcomes. To understand the calculations of partial expectation, consider a binomial model in which the strike price is \$70, and the stock price at expiration can be \$20, \$40, \$60, or \$80, with probabilities $1/8$, $3/8$, $3/8$, and $1/8$. If a put is in the money at expiration, the stock price is either \$20 or \$40. Suppose that for these two values we sum the stock price times the probability. We obtain

$$\sum_{S_1 < 50} \text{PROB}(S_t) \times S_t = \left(\frac{1}{8} \times \$20\right) + \left(\frac{3}{8} \times \$40\right) + \left(\frac{3}{8} \times \$60\right) = \$40$$

The value \$40 is clearly not an expected stock price since it is below the lowest possible price (\$20). We call \$17.50 the **partial expectation** of the stock price conditional upon $S_t < (\$70)$.

11. Participating Swap

Allows the fixed rate to be adjusted downward during the life of the swap, depending on the rate for payments indexed to a long-term rate.

12. Partnership

A partnership brings two or more individuals together to invest their time, energy, and talents in the firm. Organizing a partnership is relatively simple, although some

legal documents may be needed to spell out the percentage ownership, rights, and duties of each partner. By drawing on the strengths of two or more individuals, each can specialize in his or her own area to help the firm achieve success. Also, the combined financial resources of two or more individuals may increase the firm's ability to raise and borrow capital.

As with a proprietorship, partnership income is taxable to each partner at his or her own personal tax rate. The partnership ends upon the death of any partner; unless other arrangements have been made, firm assets may need to be sold to settle the deceased partner's estate.

As the partners presumably manage their firm, agency costs can be zero, as long as they agree on the firm's goals, work together amicably, and trust and respect each other as professionals. Should intractable differences of opinion or suspicions arise, arguments, and even court battles, can result.

Partnerships suffer from other drawbacks. As with a proprietorship, it is difficult to value and transfer ownership in a partnership. In addition, partners are *jointly and severally liable* for the debts of the partnership. That means each partner may have to pay more than his or her proportional ownership share to settle the firm's debts in case of failure. Each partner has unlimited liability. Anyone thinking of joining a partnership should seriously consider this risk.

The liability risks just noted describe a *general* partnership. A limited partnership addresses the liability concern by identifying at least one general partner as having unlimited liability; the remaining *limited partners* face liability limited to their investment in the firm – in other words, their personal assets cannot be demanded to settle the firm's debts. However, they also are limited in that they cannot participate in the operations of the firm. Operating decisions may be made only by the general partners.

13. Passbook Savings

Nonnegotiable, small savings account evidenced by a passbook listing the account terms.

14. Passive Investment Strategy

[See **Passive management**]

15. Passive Management

Buying a well-diversified portfolio to represent a broad-based market index without attempting to search out mispriced securities.

16. Passive Portfolio

A market index portfolio. [See **Passive portfolio management**]

17. Passive Portfolio Management

An investment policy whereby managers make predetermined securities purchases regardless of the level of interest rates and specific rate expectation. Examples include following a laddered maturity strategy whereby a bank continuously buys 10-year securities as previously owned securities mature.

18. Pass-Through

[See **Mortgage-back security**]

19. Pass-Through Security

Pools of loans (such as home mortgage loans) sold in one package. Owners of pass-throughs receive all principal and interest payments made by the borrowers.

20. Past-Due Loan

A loan with a promised principal and/or interest payment that has not been made by the scheduled payment data.

21. Path-Dependent

A derivative where the final payoff depends upon the path taken by the stock price, instead of just the final stock price.

22. Path-Dependent Option

An option whose payoff depends on the whole path followed by the underlying variable-not just its final value. An Asian option is an example of path-dependent option, since Asian option that has a payoff that is based on the average price over some period of time. [See **Asian option**]

23. Payable Through Drafts

Payable through drafts resemble checks; they are written orders to pay and have the physical appearance of checks. However, they are drawn directly against the issuing firm instead of a bank. The bank receives a draft first; it sends the draft to the issuing firm and awaits approval. The bank releases funds only when the corporate issuer approves specific drafts for payment. In practice, the bank generally withholds payment for one business day and then covers the payment automatically unless directed otherwise. The issuing firm generally inspects the drafts for inaccuracies in signatures, amounts, and dates, and quickly cancels payments on issued drafts with discrepancies.

Although drafts may increase disbursement float, their main advantage lies in ensuring effective control over payments. Draft payments are popular in the insurance industry, for instance, where they allow field agents to settle claims quickly even though they lack the authority to issue checks. Drafts give the central office the flexibility to improve efficiency in field operations, yet still retain the option to block any payments deemed inappropriate.

24. Payback Method

The payback method calculates a project's payback period as a measure of how long it takes the project to pay for itself. More formally, it is the time necessary for a project to generate cash flows sufficient to recover its cost. Projects with payback periods less than a management-determined cutoff are acceptable. Projects with longer paybacks are rejected.

The payback method has none of the characteristics we want from a project selection method. First, it ignores the time value of money, summing periodic cash flows without regard for the differences in the present values of those dollars. Second, the payback method fails to account for all relevant cash flows, ignoring those that accrue after the payback period. Third, the payback period gives no indication of the absolute change in shareholder wealth due to a particular project. Finally, the decision criterion is quite subjective. The determination of an appropriate payback period is based solely upon management's opinions and perceived needs. It has no relationship to the project's required return.

Some firms use a discounted payback method, in which the payback is computed using the present value of the cash inflows. [See **Discount payback rule**]

25. Payback Period Rule

An investment decision rule which states that all investment projects that have payback periods equal to or less than a particular cutoff period are accepted, and all of those that pay off in more than the particular cutoff period are rejected. The payback period is the number of years required for a firm to recover its initial investment required by a project from the cash flow it generates.

26. Payer Swaption

A swaption giving the holder the right to be the fixed-rate (or fixed price) payer in the swap.

27. Paylater Strategy

Generally used to refer to option strategies in which the position buyer makes no payments unless the option moves more into the money. This is an exotic option in which the premium is paid only at expiration and only if the option is in the money.

28. Payment Date

The firm mails checks to shareholders on the payment date. [See also **Dividend declaration date**]

29. Payment-in-Kind

Payment-in-kind (PIK) bonds often are issued by cash-strapped firms and firms doing leveraged buyouts. The PIK provision allows the issuer to pay coupon interest in the early years of the issue in the form of either cash or bonds with values equal to the coupon payment. Such bonds help reduce the issuer's cash outflows, but at a cost of increasing the debt. Investors also assume a risky position; unless the issuer's cash situation improves, they find themselves increasing their exposure to the questionable lender.

30. Payments Pattern Approach

Describes the lagged collection pattern of receivables. For instance the probability that a 72-day-old account will still be unpaid when it is 73 days old. [See **Receivable balance pattern**]

31. Payoff

The cash realized by the holder of an option or other derivatives at the end of its life.

32. Payoff Diagram

A graph in which the value of a derivative or other claim at a point in time is plotted against the price of the underlying asset.

33. Payout Ratio

Proportion of net income paid out in cash dividends.

34. Payout Phase

The payout phase usually starts at retirement, when the investor typically has several options, including the following:

1. Taking the market value of the shares in a lump sum payment.
2. Receiving a fixed annuity until death.
3. Receiving a variable amount of money each period that is computed according to a certain procedure.

35. Payout Policy

Payout policy can be measured by the payout ratio. The optimum payout ratio is generally dependent on the amount of internal funds needed. In addition, investor's preference on cash dividend payment can also influence the optimum payout ratio.

36. Peak

The transition from the end of an expansion to the start of a contraction in business cycle.

37. Peak Exposure

For market-driven instruments, the maximum (perhaps netted) exposure expected with 95% confidence for the remaining life of a transaction. CreditMetrics does not utilize this figure because it is not possible to aggregate tail statistics across a portfolio, since it is not the case that these "peaks" will all occur at the same time. [See **CreditMetrics**]

38. P/E Effect

Fundamental analysis calls on much wider range information to create portfolios than does technical analysis. One of the criteria is to use P/E ratio information to formulate portfolios. It has been found that portfolios of low P/E stocks have exhibited higher average risk-adjusted returns than high P/E stocks.

39. P/E Ratio

A firm's stock price per share divided by earnings per share.

40. Pecking Order Hypothesis

The pecking order hypothesis is a perspective based upon repeated observations of how corporations seem to raise funds over time. The theory behind this perspective was developed from the **information asymmetry** problem – namely, that management knows more about the firm and its opportunities than the financial marketplace does, and that management does not want to be forced to issue equity when stock prices are depressed. [See also **Information asymmetry**]

Evidence shows that corporations mainly rely on internal funds, especially new additions to retained earnings, to finance capital budgeting projects. If they need outside financing, firms typically issue debt first, as it poses lower risk on the investor than equity and lower cost on the corporation. Should a firm approach its debt capacity, it may well favor hybrid securities, such as convertible bonds, over common stock. As a last resort, the firm will issue common equity. Thus, firms have

a financing “pecking order,” rather than a goal to maintain a specific target debt-to-equity ratio over time.

Under this pecking order hypothesis, financial theory has come full circle. Like Modigliani and Miller’s original work, the pecking order hypothesis implies that firms have no optimal debt-to-equity ratios. Instead, they follow the pecking order, exhausting internal equity (retained earnings) first and resorting to external equity (new issues of common stock) last. Observed debt ratios represent nothing more than the cumulative result of a firm’s need to use external financing over time.

Under the pecking order hypothesis, firms with high profitability should have *lower* debt ratios, as these firms’ additions to retained earnings reduce their need to borrow. Under the static trade-off hypothesis, a firm with high profitability ratios should have a lower probability of bankruptcy and a higher tax rate, thus leading to *higher* debt ratios. Most empirical evidence resolves this conflict in favor of the pecking order hypothesis; studies find that more profitable firms tend to have lower debt ratios.

What if the pecking order hypothesis is correct and the firm has no optimal capital structure? Recall that the cost of capital represents the minimum required return on capital budgeting projects. Management must determine the firm’s cost of capital regardless of personal beliefs about the existence of an optimal capital structure. Target capital structure weights should reflect management’s impression of a capital structure that is sustainable in the long run and that allows for financing flexibility over time. Should a firm fail to earn its cost of capital, shareholder wealth will decline.

The debate over optimal capital structure is not resolved. Empirical studies and surveys of corporate practice have supported both the static trade-off and the pecking order theories. Part of the uncertainty over which perspective is correct comes from blends between capital structure choices that depart from “plain vanilla” debt and equity. In recent decades, firms have devised myriad financing flavors. Consequently, many firms have several layers of debt and several layers of equity on their balance sheets. Debt can be made convertible to equity; its maturity can be extended, or shortened, at the firm’s options; debt issues can be made senior or subordinate to other debt issues. Likewise, equity variations exist. Preferred equity has gained popularity since it increases a firm’s equity without diluting the ownership and control of the common shareholders; it also increases future financing flexibility by expanding the firm’s capacity for debt issues. Firms can have different classes of common equity, providing holders with differing levels of dividend income or voting rights.

In sum, pecking order in long-term financing is a hierarchy of long-term financing strategies, in which using internally generated cash is at the top and issuing new equity is at the bottom.

41. Peer Group

Sample firms used to generate average reference data for comparison with an individual firm’s performance data.

42. PEG Ratio

PEG ratio is defined as price-earnings ratio divided by earnings growth rate. The price/earnings to growth (PEG) ratio is used to determine a stock's value while taking the company's earnings growth into account and is considered to provide a more complete picture than the PE ratio. While a high PE ratio may make a stock look like a good buy, factoring in the company's growth rate to get the stock's PEG ratio can tell a different story. The lower the PEG ratio, the more the stock may be undervalued given its earnings performance. The PEG ratio that indicates an over or underpriced stock varies by industry and by company type, though a broad rule of thumb is that a PEG ratio below one is desirable. Also, the accuracy of the PEG ratio depends on the inputs used. Sustainable growth rate is usually used to estimate earnings growth rate. In Appendix 2B, we introduce two possible methods to calculate it. However, using historical growth rates, for example, may provide an inaccurate PEG ratio if future growth rates are expected to deviate from historical growth rates. To distinguish between calculation methods using future growth and historical growth, the terms "forward PEG" and "trailing PEG" are sometimes used.

43. Peggers

Nations that strive to keep the exchange value of their currencies within a fixed range around the value of some other currency or basket of currencies.

44. Pension Benefit Guarantee Corporation (PBGC)

The Employees Retirement Income Security Act (ERISA) of 1974 established the PBGC, which is a government-run insurance system that ensures that employees of companies that go bankrupt will receive their pension benefits.

45. Pension Funds

Financial service firms selling retirement plans to their customers in which savings are set aside in accounts established in the customers' names and allowed to accumulate at interest until those customers reach retirement age.

46. Percentage of Sales Method

The percentage of sales method is a more complex financial planning model than the internal growth, sustainable growth, or external financing needs models. The percentage of sales method generates a set of *pro forma* or *forecasted balance sheets and income statements* for the firm. The analyst projects what will happen to the firm's accounts over time, which supports an estimate of the firm's external financing needs for a particular period.

The first step of the percentage of sales method is implied by the method's name. Using historical data, the analyst divides each balance sheet and income statement

item by sales revenue. The resulting ratios are examined to see which accounts have maintained fairly constant relationships or trends with respect to sales.

The second step of the percentage of sales method is to estimate future sales levels. This estimate can rely on market research studies or on an analysis of internal or sustainable growth rates.

In the third step the analyst can construct projected financial statements. This process begins by placing the sales forecast at the top of the income statement. To forecast the value of income statement items having a steady or predictable relationship to sales, the analyst assumes this relationship will continue. For items that do not have a consistent relationship to sales, other assumptions will be needed to forecast their values. For example, current credit market conditions may suggest holding interest expense constant or projecting it to grow at a predetermined rate; projected taxes will reflect the firm's tax rate.

Similarly, the analyst projects balance sheet accounts based upon their relationship with sales revenue. Accounts that lack consistent relationships to sales may be assumed to be held constant or to change in a manner consistent with recent trends and future market projections.

The analyst estimates retained earnings by adding the forecasted addition to retained earnings to the existing retained earnings balance. The forecasted addition to retained earnings is the net income on the pro forma (projected) income statement less any dividends, that is:

$$\text{Projected retained earnings} = \text{Existing retained earnings} + \text{Projected net income} \\ - \text{Estimated dividend payment}$$

The accounting identity requires that total assets equal total liabilities and equity; in the first pass, however, the percentage of sales method will rarely produce this equality. To balance the pro forma balance sheet, the analyst inserts a plug figure, so that:

$$\text{Total assets} = \text{Total liabilities} + \text{Stockholders' equity} + \text{Plug}$$

The plug figure, sometimes labeled "external funds needed" or "external funds required," typically represents an addition to or subtraction from notes payable to restore equality to the balance sheet equation. A positive plug figure suggests that additional short-term borrowing will be needed to finance the firm's growth plans. (Of course, this need for funds also can be met by issuing long-term debt or equity.) A negative plug figure suggests that project operating results will generate excess cash which the firm can use to reduce its short-term or long-term borrowing or to repurchase stock.

47. Percentage of Sales Method for Finance

The percentage of sales method is used to develop a budgeted set of financial statements. Each historical expense is converted into a percentage of net sales, and

these percentages are then applied to the forecasted sales level in the budget period. The approach can also be used to forecast some balance sheet items, such as accounts receivable, accounts payable, and inventory.

48. Percentage of Sales Method for Financial Planning and Forecasting

Following Lee et al. (2022), the procedure of using percentage sales methods for financial planning and forecasting can be briefly described as follows:

1. Calculate by dividing company sales into both balance sheet and income statement to obtain standardized financial statements
2. Forecast growth rate and outside financing
3. Forecast external financing needs
4. Combine information from steps 1, 2, 3 to obtain pro forma financial statements
5. Use pro forma financial statements to perform future financial planning and forecasting

Based upon percentage sales method, Warren and Shelton (1971) have developed a 20-equation simultaneous model for financial planning analysis and forecasting.

49. Percentile Level

A measure of risk based on the specified confidence level of the portfolio value distribution: for example, the likelihood that the portfolio market falls below the 99th percentile number is 1%.

50. Perfect Markets

Perfectly competitive financial markets.

51. Perfectly Competitive Financial Markets

Markets in which no trader has power to change the price of goods or services. Perfect markets are characterized by the following conditions: (1) Trading is costless, and access to the financial markets is free. (2) Information about borrowing and lending opportunities is freely available. (3) There are many traders, and no single trader can have a significant impact of market prices.

52. Performance Shares

Shares of stock given to managers on the basis of performance as measured by earnings per share and similar criteria – a control device used by shareholders to tie management to the self-interest of shareholders.

53. Permanent Working Capital

Some working capital needs persist over time, regardless of seasonal or cyclical variations in sales. The firm will always maintain some minimum level of cash, accounts receivable, or inventory; this is permanent working capital and is usually some target percentage of sales.

54. Perpetual Option

An option that never expires.

55. Perpetual Preferred Stock

Nonmaturing preferred stock.

56. Perpetuity

A constant stream of cash flows without end. A British consol is an example. Consider a consol that pays a coupon of C dollars each year and will do so forever. Simply applying the present value (PV) formula gives us $PV = \frac{C}{r}$.

57. Personal Banker

Individual assigned to a bank customer to handle a broad range of financial services.

58. Personal Trust

An interest in an asset held by a trustee for the benefit of another person.

59. Perquisites

Management amenities such as a big office, a company car, or expense-account meals. “Perks” are agency costs of equity, because managers of the firm are agents of the stockholders.

60. Pie Model of Capital Structure

A model of the debt-equity ratio of the firms, graphically depicted in slices of a pie that represents the value of the firm in the capital markets.

61. Plain Vanilla

A term used to describe a standard deal. The most basic type of interest rate swap is known as a “plain vanilla swap.”

62. Planned Amortization Class

A collateralized mortgage obligation (CMO) that receives principal from the underlying mortgages based on a predetermined payment schedule, where the payments vary depending on whether prepayments fall inside or outside some predetermined range.

63. Planning Horizon

Planning horizon can be classified into short-term and long-term horizon. Short-term horizon is less than 1 year, while long-term horizon represents longer than 1 year.

64. Planning Phase of Capital Budgeting

The planning or identification phase examines areas of opportunity or change that could offer profitable investment.

Over time, managers define and redefine the firm's mission, or "vision" and the strategies they will use to accomplish that mission. This long-term plan provides a foundation for the following 5–10 years of operation planning for the firm. The long-term plan is operationalized, or implemented, in the annual capital budget. To develop the capital budget, managers must find investment opportunities that fit within the overall strategic objectives of the firm. In addition, they must consider the firm's position within the various markets it serves and the likely plans of its competitors. Attractive capital budgeting projects are those that take the firm from its present position to a desired future market position. Two popular and well-known methods that managers use to identify potentially attractive capital budgeting projects are the **business strategy matrix** and **SWOT analysis**. [See also **Business strategy matrix** and **SWOT analysis**]

65. Pledged Securities

Bank securities (either treasury or municipal securities) pledged as collateral against deposit liabilities such as Treasury deposits, municipal deposits, and borrowing from Federal Reserve banks. These pledged securities are often held by a third party trustee and cannot be sold without a release.

66. Pledging

In pledging, the firm offers its receivables as security for a cash advance. The lender who accepts and discounts the receivables may be a commercial bank or a specialized industrial finance company.

The first step in setting up a pledging relationship is to negotiate a formal agreement between the borrower and the lender. Once the agreement has been reached and a legal contract signed, the borrower can begin to present its receivables.

The lender gives the borrower the face value of the invoices less its own charges. That is, the lender buys the invoices at a discount, paying less than the amount it hopes to collect.

Almost all pledging agreements have two important provisions: the lender's right to recourse, and its right to reject invoices. In the event that the customer defaults and fails to pay the sum invoiced, the borrower is obligated to assume responsibility for the outstanding amount.

The lender also has the right to select only those invoices that it will finance and reject those it considers too risky. It is estimated that the rejection rate could reach as high as 50%.

Pledging, or discounting, receivables is not a cheap source of credit. During most of the 1980s, when the commercial bank lending rate varied between 8% and 15%, the cost of discounting was about 20%. Similar rate differentials exist today. In addition, the lender often charges yet another fee to cover its expenses to appraise credit risks. Consequently, this source of short-term financing is used mostly by companies that have no other source of funds open to them, primarily smaller companies. For such companies, however, this offers two advantages. First, after the initial agreement has been reached, the method is fairly informal and automatic, except for the rejection of invoices for bad risk. Second, the customer being invoiced receives no information that the borrowing company is in financial trouble; he or she simply sends in a check in the normal way and never knows that it has been assigned to a third party. For this reason, pledging receivables is sometimes called **nonnotification financing**.

67. Plug

A variable that handles financial plan. [See **Percentage of sales method**]

68. Plowback Ratio

The proportion of the firm's earnings that is reinvested in the business (and not paid out as dividends). The plowback ratio equals 1 minus the dividend payout ratio.

69. Point

Mortgage lenders customarily charge initial service fees, known as points, at the time of the loan origination. A point is 1% of the principle of the loan.

70. Poison Pill

Strategy by a takeover target company to make a stock less appealing to a company that wishes to acquire it. Examples of such delaying tactics, proxy defenses, or poison pills include:

1. Provisions that require super-majorities (e.g., two-thirds) of existing shareholders to approve any takeover

2. The decision to place some, rather than all, board seats up for election every year, thus delaying the ability of an acquirer to control the firm
3. Provisions to allow the board to authorize and issue large quantities of stock or to repurchase outstanding bonds in the event of a takeover attempt
4. Provisions that stipulate expensive payouts to existing managers in the face of any successful buyout
5. The establishment of advance notice requirements, so shareholders must meet deadlines for presenting business or director nominations at shareholder meetings
6. Restrictions on the ability of shareholders to call special meetings

71. Political Risk

Investors in nondomestic securities face a number of risks beyond those of domestic securities. Political risk can affect a bond investor in a number of ways. A foreign government may block currency exchanges, preventing the investor from repatriating coupon income.

72. PO

Principal Only. A mortgage-backed security where the holder receives only principal cash flows on the underlying mortgage pool.

73. Point of Sale

Electronic terminals that enable customers to directly access deposit accounts.

74. Poisson Distribution

A probability distribution that counts the number of events occurring in an interval of time, assuming that the occurrence of events is independent.

75. Poisson Process

A process describing a situation where events happen at random. The probability of an event in time (Δ_t) is $(\lambda\Delta_t)$, where (λ) is the rate (intensity) of the process.

76. Pooling of Interests

The general idea motivating the pooling treatment is that the business combination was not a purchase-sale transaction but rather a combining of interests. Hence, the prior accounting valuations are maintained and merely added together for the combined firm. Moreover, from an accounting standpoint, the two firms are considered to have been joined from day 1 and the accounting reports are restated as if they had been joined.

77. Portfolio Analysis

A portfolio is any combination of assets or investments. A firm can be considered a portfolio of capital budgeting projects.

Expected Return on a Portfolio

The expected rate of return on a portfolio, $E(R_p)$, is simply the weighted average of the expected returns, $E(R_i)$, of the individual assets in the portfolio:

$$E(R_p) = \sum_{i=1}^n w_i E(R_i)$$

where w_i is the weight of the i th asset, or the proportion of the portfolio invested in that asset. The sum of these weights must equal 1.0.

$E(R_i)$ is used to stand for the expected return on a risky asset. Whenever risk exists, the actual return is not known beforehand. We know that there is an asset which, for all intents and purposes, is considered risk-free: the Treasury bill or T-bill. Let R_f denote the nominal return on a risk-free asset. Since it has no risk, the expected nominal T-bill return is the same as its actual return.

Variance and Standard Deviation of Return on a Portfolio

The total risk of a portfolio can be measured by its variance or the standard deviation of its returns. Lower portfolio variability arises from the benefits of **diversification**. [See also **Diversification**] The benefits of diversification are greatest when asset returns are strongly negatively correlated, that is, when they tend to move in opposite directions over time.

Portfolio variance is affected not only by the variance of each asset's return but also by **covariances** between returns. [See also **Covariance**] The variance of a two-asset portfolio is computed by summing the squared weights of each asset times the asset's variance and then adding a term to capture the covariance of the two assets:

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \text{cov}(R_1, R_2)$$

where W_1 and W_2 are weights associated with first and second security respectively. σ_1^2 σ_2^2 are variance for first and second security respectively. $\text{Cov}(R_1, R_2)$ represents covariance between R_1 and R_2 .

We can also express the portfolio variance in terms of the **correlation** coefficient [See also **Correlation**]:

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \rho_{12} \sigma_1 \sigma_2$$

where (ρ_{12}) represents the correlation coefficient between R_1 and R_2 .

The standard deviation of the portfolio's returns is simply the square root of this variance.

78. Portfolio Cushion

In general, portfolio insurance can be thought of as holding two portfolios, the first portfolio can be viewed as the safe or riskless portfolio with value equal to the

level of protection desired. This level is called the floor and is the lowest value the portfolio can have. For certain strategies this can be held constant or allowed to change over time as market conditions or needs change. The second portfolio consists of the difference between the total value of the portfolio and the floor, commonly called the portfolio cushion. These assets consist of a leveraged position in risky assets. To insure the portfolio, the cushion should be managed as never to fall below zero in value because of the limited-liability property of common stock.

79. Portfolio Immunization

Making a portfolio relatively insensitive to interest rates.

80. Portfolio Insurance

The practice of using options or dynamic hedge strategies to provide protection against investment losses while maintaining upside potential. In addition, it can use an appropriate mix of treasury bills and security to create a payoff pattern identical to the pattern of an option on the underlying security. This kind of artificial option can be used to perform portfolio insurance (see Rubinstein 1985, pp. 42–52 for details).

81. Portfolio Management

Process of combining securities in a portfolio tailored to the investor's preferences and needs, monitoring that portfolio, and evaluating its performance.

82. Portfolio Opportunity Set

The expected return-standard deviation pairs of all portfolios that can be constructed from a given set of assets.

83. Position Limit

The maximum position a trader (or group of traders acting together) is allowed to hold.

84. Positive Covenant

Part of the indenture or loan agreement that specifies an action that the company must abide by.

85. Positive Float

The firm's bank cash is greater than its book cash until the check's presentation.

86. Post

Particular place on the floor of an exchange where transactions in stocks listed on the exchange occur.

87. Post Audit

The major proportion of control phase for capital budgeting process is the post audit of the project, through which past decisions are evaluated for the benefit of future capital expenditure.

88. Post-Earnings-Announcement Drift (PEAD)

In financial economics and accounting research, post-earnings-announcement drift or PEAD (also named the SUE effect) is the tendency for a stock's cumulative abnormal returns to drift in the direction of an earnings surprise for several weeks (even several months) following an earnings announcement.

89. Power Option

An option where the payoff is based on the price of an asset raised to a power. For example, a power option for call can be defined as $\text{Max}(S^b - K^b, 0)$, where S and K are stock price per share and exercise price per share respectively; b is the power.

90. Preauthorized Check System

A preauthorized check (PAC) system is a type of cash collection arrangement that may be more useful to firms such as insurance, finance, leasing, and mortgage companies. The PAC is a commercial instrument that is used to regularly transfer funds between demand deposit accounts. Through such a preauthorized indemnification agreement, the collecting firm is authorized to draw a check at specified intervals and in specified amounts on the customer's demand deposit account. An example is a monthly mortgage payment. The PAC reduces mail, processing, and collection float and ensures that the company gets its money by a specified date.

91. Preferred Habitat Theory

Investors prefer specific maturity ranges but can be induced to switch if premiums are sufficient. In other words, markets are not so segmented that an appropriate premium cannot attract an investor who prefers one bond maturity to consider a different one. [See **Market segmentation theory**]

92. Preferred Stock

A type of stock whose holders are given certain priority over common stockholders in the payment of dividends. Usually the dividend rate is fixed at the time of issue. Preferred stockholders normally do not receive voting rights.

93. Premium

The forward rate either will be at a discount or a premium to the spot rate. A currency is selling at a premium if it can purchase more units of foreign currency in the forward market.

94. Premium Bonds

When a bond's price exceeds its par value, it is said to be selling at a *premium*, and it is called a premium bond. In most cases where the bond sells at a premium, interest rates have fallen after the bond's issue.

The price of a premium bond will fall as it nears maturity if the market rate remains the same, since at maturity its price will equal its par value.

95. Premium on a Bond

Difference between the price of a bond and its par value when the price is higher. When the price is lower than the par value, then this difference is the discount on a bond.

96. Prepaid Forward Contract

A forward contract calling for payment today and delivery of the asset or commodity at a time in the future.

97. Prepaid Forward Price

The price the buyer pays today for a prepaid forward contract.

98. Prepaid Swap

A swap contract calling for payment today and delivery of the asset or commodity at multiple specified times in the future.

99. Prepayment Function and Model

A function estimating the prepayment of principal on a portfolio of mortgages in terms of other variables. Refinancing and housing turnover are two principal sources

for prepayment. There are several prepayment models to estimate the rate of prepayment. The most well-known model is the model developed by public security association.

100. Prepayment Speed

The percentage of the outstanding principal that is prepaid above and beyond normal amortization. [See **Prepayment function and model**]

101. Prepayment Penalties

Prepayment penalties, which lender charges borrower for his (or her) prepayment on mortgage, are designed to compensate for the uncertainty in asset management caused by a prepayment. Lenders face potentially large volumes of prepayments if market yields fall and borrowers with fixed mortgage rates refinance their homes at lower rates.

102. Present Value

The present value of a cash flow is the amount which, if it were invested today at r percent per year for n years, would grow to equal the future cash flow. The present value (PV) represents the maximum price we are willing to pay today in order to receive the future cash flow, FV.

To solve for PV, the present value, we obtain:

$$PV = FV_n X \left[\frac{1}{(1 + r)^n} \right]$$

Like future values, present values are additive as long as the present values occur at the same point in time. If a problem involves several future cash flows, one can easily find their total present value simply by adding the individual present values at time zero.

103. Present Value Factor

Factor used to calculate an estimate of the present value of an amount to be received in a future period. [See **Present value**]

104. Price Elasticity

The ratio of changes in the price of a debt security to changes in its yield.

105. Price Participation

The extent to which an equity-linked note benefits from an increase in the price of the stock of index to which it is linked.

106. Price of Credit

The rate of interest that must be paid to secure the use of borrowed funds.

107. Price Risk

It is one of the components of the interest rate risk, and another component is the coupon-reinvestment risk. Price risk occurs if interest rate change before the target date and the bond is sold prior to maturity. At that time the market price will differ from the value at the time of purchase. If rate increase after the purchase date, the price the bond would be sold at would be below that had been anticipated. If the rate decline, the realized price would be above what had been expected. Increase in interest rates will reduce the market value of a bond below its par value. But it will increase the return from the reinvestment of the coupon interest payment. Conversely, decrease in interest will increase the market value of a bond above its par value but decrease the return on the reinvestment of the coupons. In order for a bond to be protected from the change in interest rate after the purchase. The price risk and coupon reinvestment must offset each other.

108. Price Takers

Individuals who responds to rates and prices by acting as though they have no influence on them.

109. Price Value of a Basis Point

The change in the value of a fixed-income asset resulting from a one basis point change in the asset's yield to maturity. One basis point represents 0.25%.

110. Price Volatility

A factor that is the single most important variable affecting the speculative value of the option – the price volatility of the underlying stock. The greater the probability of significant change in the price of the stock, the most likely it is that the option can be exercised at a profit before expiration.

111. Price/Earnings Ratio (P/E ratio)

The ratio of a stock's price to its earnings per share. Also referred to as the P/E multiple. The P/E ratio tells us how much stock purchasers must pay per dollar of earnings that the firm generates. [See **Market value ratios**]

112. Price-to-Book-Value Ratio

[See **Market value ratios**]

113. Price-Variable Cost Margin

A factor affecting business risk is the firm's ability to maintain a constant, positive difference between price and per-unit variable cost:

$$\text{Margin} = \frac{\text{Price per unit} - \text{Variable cost per unit}}{\text{Price}}$$

This is one of the factors used to determine the business risk. [See **Business risk**]

114. Price-Weighted Index

In a price-weighted index the basic approach to sum the prices of the component securities used in the index and divide this sum by the number of components; in other words, to compute a simple arithmetic average. The Dow-Jones Industrial Average (DJIA) is the most familiar index of this type. To allow for the impact of stock splits and stock dividends, which could destroy the consistency and comparability of price-weighted index data over time, an adjustment of either the reported price data or the divisor itself is required.

115. Pricing Grid

A schedule of credit spreads listed by credit rating that is applied to either a loan or Credit-Sensitive Note (CSN) upon an up (down) grade of the obligor of issuer. If the spreads are specified at market level, then such terms reduce the volatility of the value across all non-default credit quality migrations by keeping the instrument close to par.

116. Primary Capital

The sum of common stock, perpetual preferred stock, surplus, undivided profits, contingency and other capital reserves, valuation reserves, mandatory convertible securities, and minority interest in consolidated subsidiaries at a bank.

117. Primary Dealers

Security firms that acquire securities from the government and other borrowers and make markets from those securities standing ready to buy and sell them at posted prices.

118. Primary Market

The primary market is the market for original securities, or first-time issues. For example, a corporation first sells its stock to the public in an initial public offering. Such a sale is a primary market transaction. If, after additional growth, the firm

determines that it needs more equity capital, it can sell another new issue of stock in the primary market. In general, whenever a firm raises money by selling shares, bonds, commercial paper, or other securities to investors, it does so in primary market transactions. Government issues of Treasury bills and bonds, as well as state and local government security issues, also occur in the primary market.

119. Primary Reserves

The most liquid asset held by a bank, consisting of its vault cash and the deposits it has places with other banks.

120. Primary Securities

The IOUs issued by borrowers from a financial intermediary and held by the intermediary as interest-bearing assets.

121. Prime Rate

The rate of interest charged by commercial banks varies in two ways: the general level of interest rates varies over time, and, at any given time, different borrowers pay different rates because of varying degrees of creditworthiness. The base rate for most commercial banks traditionally has been the prime rate, although in times of soaring market interest rates, some of the larger banks experiment with marginal pricing schemes. The prime rate is the rate that commercial banks charge their most creditworthy business customers for short-term borrowing. The financial press splashes news of any change in this rate across the front page. Congress and the business community speculate about the prime's influence on economic activity, because it is the baseline rate for loan pricing in most loan agreements.

In the latter part of 1971, a large, money-center bank instituted a floating prime rate linked by a formula to the market-determined commercial paper rate. The formula required weekly reviews of the prime rate, with adjustments in minimum steps of one-eighth of a percentage point. The formula kept the prime approximately 50 basis points above the average rate on 90-day commercial paper placed through dealers. The choice of the commercial paper rate reflected the ease of substituting short-term bank loans for commercial paper. Historically, the prime has served as a base line for loan pricing; a loan contract might state its interest rate as "prime plus two" or "120% of prime."

However, as the banking industry has begun to price its loans and services more aggressively, the prime rate has become less important. As the use of the prime rate has declined, compensating balances have become less popular, as well. The current trend is to price a loan at a rate above the bank's marginal cost of funds, which typically is reflected by the interest rate on a certificate of deposit. The bank adds an interest-rate margin to this cost of funds, and the sum becomes the rate it charges the borrower. This rate changes daily, in line with the bank's money market rates.

122. Primitive Security, Derivative Security

A *primitive security* is an instrument such as a stock or bond for which payments depend only on the financial status of its issuer. A *derivative security* is created from the set of primitive securities to yield returns that depend on factors beyond the characteristics of the issuer and that may be related to prices of other assets.

123. Principal

The value of a bond that must be repaid at maturity. Also called the face value or the par value.

124. Principal-Agent Problem

The principals, or owners of the firm hire agents, or managers, to run the firm in the best interests of the principals. But ethical lapses, self-interest, or the owners' lack of trust in the managers can lead to conflicts of interest and suspicions between the two parties. This problem in corporate governance is called the principal-agent problem.

The shareholders of a firm elect a board of directors. In theory, the board's role is to oversee managers and ensure that they are working in the best interests of the shareholders. In practice, however, the board often has a closer relationship with management than with the shareholders. For example, it is not unusual for the firm's top executives to sit on the firm's board of directors, and the firm's top executives often nominate candidates for board seats. These relationships can obscure loyalties and make the board a toothless watchdog for shareholders' interest.

Managers, acting as agents, may pursue their own self-interest by increasing their salaries, the size of their staffs, or their perquisites (better known as "perks"), which might include club memberships and the use of company planes or luxurious company cars. Management, in conjunction with the board, may seek to fend off takeovers that would allow shareholders to sell their shares at a price above the current market price, or they may try to preempt such merger or acquisition attempts by seeking changes in the corporate charter that would make such takeovers difficult to pursue.

Other examples of principal-agent relationships that one may relate to: voters (principals) elect officeholder (agents) to work in the best interest of the public; but political action committee (PAC) contributions to political campaigns may affect politician's actions if elected. Investors (principals) trust the advice of stockbrokers (agents) when investing their savings; but many stockbrokers earn their paycheck by generating commissions on trading. Accountants and lawyers (agents) often bill their clients (principals) by the number of hours they work, irrespective of whether the client's tax bill was minimized or the court case was won.

125. Principal Components Analysis

A multivariate analysis aimed at finding a small number of factors that describe most of the variation in a large number of correlated variables. (Similar to a factor analysis.)

126. Principle of Diversification

Highly diversified portfolios will have negligible unsystematic risks. In other words, unsystematic risks disappear in portfolios, and only systematic risks survive.

127. Private (or Direct) Sale

Placing securities with one or a limited number of investors rather than trying to sell them in the open market.

128. Private Pass-Throughs

Securities issued against a group of conventional or guaranteed home mortgages with the interest and principal payments on the securities derived from the cash flow generated by the home mortgages.

129. Private Placement

Firms in the *Fortune* 500 obtain over one-half of their long-term debt from private sources. These private sources include loans from banks and finance companies, as well as private placements of debt. If the largest US firms with access to the public debt market do most of their long-term borrowing in the private market, smaller public and private firms rely on privately arranged loans even more heavily.

A private placement or sale of debt is similar to a private placement of equity. The borrower and lenders negotiate the terms of the placement: the amount of the loan, its interest rate, the timing of cash flows, lender security, and covenants. An investment bank may act as a broker to help place the private debt with accredited investors (those who meet SEC rules regarding net worth and investment experience). Most privately placed debt matures in 5–20 years and pays fixed interest rates.

Large insurance companies and pension funds are major purchasers of private debt. These lenders typically have long investment horizons and low liquidity needs, so they are ideal private placement investors.

By avoiding the need to register securities with the SEC, a borrower can save on some of the up-front expenses of issuing debt securities. Lack of registration, however, makes private placements less liquid than publicly issued bonds. Some market participants have attempted to increase the liquidity of the private placement market, such as NASDAQs PORTAL (Private Offering, Resale, and Trading through Automated Linkage) system. Even with such trading, private placements can be bought and sold only among accredited investors. Due to the lack of public

disclosure, the investing public is not allowed under SEC rules to trade or invest in private placements.

At year-end 1991, almost every one of the industrial firms with investment-grade S&P bond ratings had total firm exceeding \$500 million. This statistic suggests that most small firms are shut out of the public capital markets because of those markets' aversion to below-investment-grade issues. Private placements play a major role in financing growth and expansion for many of these small- and medium-sized firms. Private placements do this without imposing excessive interest expenses. The effect of liquidity risk appears to be reduced by a number of different factors, including long investment horizons and little need for immediate liquidity by the investors, such as life insurers; the freer flow of information that occurs during negotiations; the ability to negotiate covenants; and the access that lenders have to firm-specific information (including discussions with top management and on-site plant visits) as they conduct their due diligence analysis. Should default occur, the ability to renegotiate terms and conditions also easier than in the case of a default on a public issue.

130. Private Placement of Equity

A private placement raises funds by allowing outside private investors to purchase shares in the firm. Such a deal may be difficult to arrange, however, as any new investor(s) may suspect the original owners' motives and question their ability to successfully invest the funds to create future value. Arrangements for private placements may be made by a business broker or an investment banker, who earns a commission for finding a qualified investor. To limit the cost and ensure the compatibility of the new owners, current shareholders also may seek additional investors among their friends, relative, and other contacts.

A private placement of equity can provide needed new capital, but only at the cost of diluting ownership. The original owners now must share control, voting rights, and company profits with additional investors. In addition, there is the problem of placing a value on the firm's privately held common stock. Private firms typically lack audited financial statements and other safeguards that reduce agency costs. Thus, new investors may resist paying what the current owners feel is a fair price for their equity. Equity investments in private firms can impose a great deal of liquidity risk, because no well-developed secondary market trades shares in firms that are not publicly owned. In recent years, the Securities and Exchange Commission (SEC) has taken some steps to increase liquidity in the private placement market. Nonetheless, a great deal of liquidity risk still remains for investors in private firms.

131. Pro Forma Financial Statements

Pro forma financial statements is a planning model that uses inputs supplied by management in the form of economic, accounting, market, and policy information. With projected financial statement data, a pro forma financial statement forecasts stock price, earnings per share, dividends per share, new equity issues, and new debt issues.

132. Probability Distribution

[See **Normal distribution**]

133. Probability of Default

[See **Default probability**]

134. Probate

Legal act of submitting a will before a court to verify authenticity of the document.

135. Problem Loans

Loan currently in default or expected to obtain default status.

136. Processing Float

[See **Float**]

137. Production Function

The production function is a technical relationship between the amount of inputs that a firm uses and the maximum level of output that can be obtained. As such, it is not based on any important behavioral assumption imposed on producers.

138. Profit

The payoff less the future value of the original cost to acquire the position. In accounting, profit refers the net income which is the last item of an income statement.

139. Profit Diagram

A graph plotting the profit on a position against a range of prices for the underlying asset. This diagram is frequently used in analyzing option strategy.

140. Profit Margin

Profits divided by total operating revenue. The net profit margin (net income divided by total operating revenue) and the gross profit margin (earnings before interest and taxes divided by the total operating revenue) reflect the firm's ability to produce a good or service at a high or low cost.

Higher profit margins generate more net income, larger additions to retained earnings, and faster growth, when all else is held constant. Should growth outpace

the planned rate, the firm can seek to finance the unexpected growth by raising its prices and/or reducing expenses in an attempt to increase its profit margin. If growth falls short of the planned rate, the firm may have to reduce prices, and therefore its profit margin, to stimulate sales. [See also **Profitability ratios**]

141. Profitability Index

A discounted cash flow technique for evaluating capital budgeting projects is the profitability index (PI), also called the *benefit/cost ratio*. The PI method computes the ration between the present values of the cash flows and initial investment:

$$PI = \frac{\text{Present value of the cash flows}}{\text{Initial cost}} = \frac{\sum_{t=1}^N \frac{CF_t}{(1+r)^t}}{I}$$

The PI measures the relative benefits of undertaking a project, namely the present value of benefits received for each dollar invested. A PI of 2, for example, means that the project returns \$2 for every \$1 invested, in present value terms. Since it would be foolish to invest in a project that returns less than a dollar for every dollar invested, the profitability index has a naturally objective decision rule: *The firm should accept a project that has a profitability index greater than 1.0 and reject a project that has a PI less than 1.0.*

The relationship between PI and NPV should be clear. Whenever NPV is positive, PI exceeds 1.0. Likewise, whenever NPV is negative, PI is less than 1.0. Thus, as with the NPV and IRR, the NPV and PI always on which projects will enhance shareholder wealth and which will diminish it. Therefore, the NPV, IRR, and PI always will agree as to whether a project should be accepted or rejected.

The profitability index considers all relevant cash flows, accounts for the time value of money, and specifies an objective decision criterion. Like IRR, however, PI measures *relative* project attractiveness; it indicates which projects add to shareholder wealth, but it gives little insight as to the amount of the change. Thus, like IRR, PI rankings of the attractiveness of mutually exclusive projects may differ from NPV rankings.

142. Profitability Ratios

Profitability ratios show the ability of a firm to use its sales, assets, and equity to generate returns. The **profit margin**, or **return on sales**, represents the proportion of each sales dollar that becomes profit or net income to the firm. The **return on assets ratio**, or ROA (sometimes called return on investment, or ROI), measures how efficiently the firm uses its total assets to generate income. Profit margin and return on assets are computed as follows:

$$\text{Profit margin} = \frac{\text{Net income}}{\text{Sales}}$$

$$\text{Return on assets} = \frac{\text{Net income}}{\text{Total assets}}$$

The return on assets ratio can be broken into two components; it equals the product of the profit margin and total asset turnover ratio:

$$\text{Return on assets (ROA)} = \text{Profit margin} \times \text{Total asset turnover}$$

$$\frac{\text{Net income}}{\text{Assets}} = \frac{\text{Net income}}{\text{Sales}} \times \frac{\text{Sales}}{\text{Total assets}}$$

This ratio gives two general strategies by which a firm can generate a high ROA. A firm can have a high profit margin with a low turnover (which is often the case for a jewelry store) or a low profit margin with a high turnover (which is often the case for a supermarket).

The **return on equity ratio (ROE)** measures profitability with respect to the stockholders' investment in the firm. It is computed as:

$$\text{ROE} = \frac{\text{Net income}}{\text{Total equity}}$$

Like return on assets, ROE can be broken down into component parts to improve insight into the means by which the firm generates income. The return on equity is identical to return on assets multiplied by the equity multiplier:

$$\frac{\text{Net income}}{\text{Total equity}} = \frac{\text{Net income}}{\text{Total assets}} \times \frac{\text{Total assets}}{\text{Total equity}}$$

Since ROA is itself comprised of two other ratios, we obtain:

$$\text{ROE} = \text{Profit margin} \times \text{Asset turnover} \times \text{Equity multiplier}$$

$$\frac{\text{Net income}}{\text{Total equity}} = \frac{\text{Net income}}{\text{Sales}} \times \frac{\text{Sales}}{\text{Total assets}} \times \frac{\text{Total assets}}{\text{Total equity}}$$

This analysis shows that a firm's return on equity may change from 1 year to the next or may differ from a competitor's ROE as a result of differences in profit margin, asset turnover, or leverage. Unlike the other measures of profitability, ROE directly reflects a firm's use of leverage, or debt. If a firm assumes more liabilities to finance assets, the equity multiplier will rise and holding other factors constant, the ROE will increase. This leveraging of a firm's return on equity does not imply greater operating efficiency, only a greater use of debt financing. Setting an optimum proportion of debt is part of the capital structure decision.

Breaking ROE into its component parts is called **Du Pont analysis**, named after the company that popularized the technique. By examining differences in the components of ROE either over time or across firms, an analyst can gain information about the strengths and weaknesses of firms. Du Pont analysis can break ROE into its

components and illustrate how the components can, in turn, be broken into their constituent parts for analysis. Thus, an indication that a firm's ROE has increased as a result of higher turnover can lead to study of the turnover ratio, using data from several years, to determine if the increase has resulted from higher sales volume, better management of assets, or some combination of the two.

It seems obvious that an analyst should prefer higher profitability ratios to lower profitability ratios. Still, the analyst must examine financial statements to determine the reasons for rising profitability and to verify that it represents truly good news about a firm. In an inflationary environment, for example, higher profitability may come from increases in sales revenues due to higher prices, while many expenses (such as FIFO inventory, depreciation, and interest expense) may be based upon historical costs. Higher profits and profitability ratios also could occur because of reductions in R&D spending or advertising expenses; such reductions may benefit the bottom line in the short run, but cutbacks in technological innovation and marketing may hurt the firm in the long run.

Changing from one generally accepted accounting principle to another also may have the effect of raising revenue, reducing expenses, and increasing profit without any real change in firm operations. Higher profits also may arise from extraordinary items, such as a successful lawsuit, or from asset sales; the analysis should remove special items from net income to obtain a clearer picture of firm profitability. The analyst always should compare several consecutive financial reports and, once again, read the financial statement footnotes to confirm that higher profitability really does represent better firm performance, and not inflation, cosmetic expense slashing, changes in GAAP, or nonrecurring items.

143. Pro Forma Financial Statements

Financial statements with projected or forecasted balance sheet and income statement data. In addition, it also includes forecasts of stock price per share, earnings per share, dividend per share, new equity issues and new debt issues. [See **Percentage of sales method**]

144. Product Differentiation

Product differentiation can generate positive net present values. Differentiation comes from consumers' belief in a difference between firms' products. Differentiation leads to an imperfect market where a firm can set prices above marginal costs, thus giving the firm some competitive advantage over its rivals. Potential sources of differentiation include advertising and promotion expenditures, marketing skills, brand loyalty, R&D, and quality differences.

145. Program Trading

Coordinated buy orders and sell orders of entire portfolios, usually with the aid of computers, often to achieve index arbitrage objectives. It encompasses several

modern investment strategies. The narrowest definition of program trading is the simultaneous placement of buy and sell orders for group of stock totaling one million or more. A common and controversial form of program trading is the simultaneous trading of stock and stock futures to profit from the change in the spread between the two sometime called index arbitrage.

146. Projected Benefit Obligation (PBO)

PBO is a measure of sponsor's pension liability that includes projected increases in salary up to the expected age of retirement.

147. Project Finance

Project finance is a technique where it is appropriate to use project-specific financing costs as required rates of return. This technique has gained popularity in recent years; it has been used to finance a variety of projects, including oil and gas development projects, R&D partnerships, and factory construction.

Project finance makes sense when a project's accounts are separated from the firm's other asset and cash flow accounts. Additionally, the project's assets must be financed by specific sources of funds whose only recourse in the case of default or project failure is to the assets of the project; in other words, the sponsoring firm is not liable for the debts of the project. Such a project also must have a definite termination time, rather than operating as a going-concern. In such cases, we can compare the project-specific financing costs to the projects returns. Returns in excess of project costs accrue to the parent firm's shareholders.

148. Promissory Note

Basically, a promissory note is an IOU in which the buyer promises to pay the seller a certain amount by the specified date for a designated order, all in writing and signed by the buyer.

149. Property-Casualty Insurance Companies

Financial service firms selling contracts to protect their customers against losses to person or property due to negligence, crime, adverse weather changes, fire, and other hazards.

150. Proprietorship

Proprietorships outnumber all other forms of business organizations in the United States. A proprietorship is simply a business owned by one person. Setting up a business is fairly simple and inexpensive – seldom more complicated than applying for a city or state license. All income is taxed as personal income to the proprietor. Depending on this person's filing status and income level, this can be an advantage

or a disadvantage. For example, depending upon the owner's level of taxable income, a proprietorship may owe more or less tax than a corporation with the same level of taxable income.

As the firm has one owner, this person's expertise determines much of the success of the firm. If additional expertise is needed, the owner must hire someone. The life of the proprietorship ends when the owner dies; in general, a proprietorship is not an asset that can be easily valued and sold.

Agency costs are nil in proprietorships, as the manager is the owner, and he or she presumably will make decisions that reflect his or her best interests. The ability to raise capital income is limited to the owner's personal wealth and credit line (although generous friends or relatives may help him or her).

Proprietorships have unlimited liability, which makes the proprietor solely responsible for all debts of the business. Should bankruptcy occur, the owner's personal assets – financial holdings, cars, house – may be forfeited to settle any debts. Losses may exceed what the proprietor has invested in the firm.

151. Prospectus

To offer stock for sale, the firm distributes a prospectus, which contains much of the same information that appears in the SEC filing. During the waiting period, the firm can distribute a **red herring** to prospective investors. [See also **Red herring**]

152. Protective Covenant

A provision specifying requirements of collateral, sinking fund, dividend policy, etc., designed to protect the interests of bondholders.

153. Protective Put

Purchase of stock combined with a put option that guarantees minimum proceeds equal to the put's exercise price.

154. Proxy

A grant of authority by the shareholder to transfer his or her voting rights to someone else.

155. Proxy Contest

Attempt to gain control of a firm by soliciting a sufficient number of stockholder votes to replace the existing management.

156. Prudent Man Rule

Requirement that a fiduciary exercise discretion, prudence, and sound judgment in managing the assets of a third party.

157. Public Confidence

The attitude of the public toward the safety and stability of financial markets and institutions.

158. Public Debt

The volume of debt obligations that are the responsibility of the federal government and therefore of its taxpayers.

159. Public Issue

Sales of securities to the public.

160. Public Offering, Private Placement

A *public offering* consists of bonds sold in the primary market to the general public; a *private placement* is sold directly to a limited number of institutional investors. [See **Private placement**]

161. Public Sale

When securities are sold in the open market to any individual or institution willing to pay the price, usually through investment bankers.

162. Publicly Traded Option

A publicly traded option is an agreement between two individuals who have no relationship with the corporation whose shares underlie the option. When a publicly traded option is exercised, money and shares are exchanged between the individuals and the corporation receives no funds. [See also **Warrant**]

163. Public Warehousing

Public warehousing, sometimes called terminal warehousing, is similar to **field warehousing**, except that the physical inventory is transferred to and stored in a warehouse operated by an independent warehousing company instead of in a segregated section of the borrower's premises. [See also **Field warehousing**] The mechanics of the financing arrangement remain the same: no inventory is released to the borrower until it repays the corresponding part of the loan. Warehouse financing is very common in the food and lumber industries. Canned goods, in particular, account for almost 20% of all public warehouse loans; however, almost any nonperishable and easily marketable commodity may be used.

164. Pull-to-Par

The reversion of a bond's price to its par value at maturity.

165. Purchase Accounting

An accounting method for acquisitions in which the assets and liabilities of the combined firm reflect a revaluation of assets and liabilities of the subject firms, thus recognizing the value of goodwill and other intangibles. [See also **Purchase method**]

166. Purchase Method

The purchase method of accounting for business combinations corresponds to the basic accounting principles for the acquisition of assets. However, in the case of business combinations, the procedure is complicated because several assets and liabilities may be acquired and more than cash may be given. Also, the excess of the price paid for the acquired asset over its book value is reflected as goodwill on the balance sheet of the acquiring firm and is amortized over a period not exceeding 40 years. Goodwill is not deductible for tax purposes, so the net result of the purchase method is a decrease in accounting earnings without the corresponding tax benefits. Hence, the purchase method is not favored by acquiring firms.

167. Purchased Call

A long position in a call. It refers to buy a call which is available in the market.

168. Purchased Put

A long position in a put. It refers to buy a put which is available in the market.

169. Purchasing Power Parity (PPP)

The purchasing power parity relates the changes in exchange rates to the relative differences in the respective rates of inflation among nations. In other words, it implies that the exchange rate adjusts to keep purchasing power constant among currency. For example, if the expected inflation rate in England is 10% and the expected inflation rate in the United States is 5%, one would expect the interest rate in England to be 5% higher than a comparable rate in the United States. Likewise, one would expect the English pound sterling to depreciate by 5% relative to the US dollar. Without these relationships, an arbitrageur could make a riskless profit by buying or selling a spot currency in the foreign exchange market, investing in the money market with the more favorable interest rate, and hedging these transactions by selling or buying the currency forward for a similar time period. This procedure,

called interest rate arbitrage, links the foreign exchange market to the money market. [See **Interest rate parity**]

170. Purchasing-Power Risk

The variability of return caused by inflation, which erodes the real value of the return. Purchasing power risk is related to the possible shrinkage in the real value of a security even though its nominal value is increasing. For example, if the nominal value of a security goes from \$100 to \$200. The owner of this security is pleased because the investment has doubled in value. But suppose that, concurrent with the value increase of 100%. The rate of inflation is 200% – that is, a basket of goods costing \$100 when the security was purchased now costs \$300. The investor has a “money illusion” of being better off in nominal terms. The investment did increase from \$100 to \$200; nevertheless, in real terms, whereas the \$100 at time zero could purchase a complete basket of goods, after the inflation only 2/3 of a basket can now be purchases. Hence, the investor has suffered a loss of value.

171. Pure Discount Bond

Bonds that pay no coupons and only pay back face value at maturity. Also referred to as “bullets” and “zeros.” [See **Discount bond**]

172. Pure Yield Pickup Swap

In a pure yield-pickup swap there is no expectation of market changes, but a simple attempt to increase yield. Basically, two bonds are examined to establish their difference in yield to maturity, with a future adjustment to consider the impact of interim reinvestment of coupons at an assumed rate of return between now and the maturity date.

173. Pure Play Method

The pure play method estimates the beta of the proposed project based on information from firms that are in similar lines of business as the project.

If the capital budgeting project involves an expansion to another country, perhaps a firm in that country will qualify as a pure play. The project’s systematic risk can be estimated by regressing the foreign firm’s stock market returns on those of a US market index. The foreign firm’s stock returns should be adjusted for exchange rate fluctuations, so exchange rate risk is included in the analysis.

The main drawback to the pure play method is that the analyst must find one or more publicly traded firms that are close proxies to the project under review. Only for publicly traded firms can the analyst find stock return data from which to estimate beta. The ideal proxy firms are single-product firms so the analysis can focus on the systematic risk of the particular project under consideration. A firm with many different product lines will complicate the comparison, as its betas will reflect the

systematic risk of the firm's overall product mix, rather than the project's line of business.

174. Put

A put is an option to sell a fixed number of shares of common stock. It is a right instead of an obligation.

175. Puttable Bonds

Puttable bonds (sometimes called **retractable bonds**) allow investors to force the issuer to redeem them prior to maturity. Indenture terms differ as to the circumstances when an investor can "put" the bond to the issuer prior to the maturity date and receive its par value. Some bond issues can be put only on certain dates. Some can be put to the issuer in case of a bond rating downgrade. Still others, nicknamed *super poison puts*, are puttable only in the case of an event such as a merger, leveraged buyout, or major financial restructuring and subsequent rating downgrade below investment quality (BBB). In any of these situations, bond investors would suffer a loss of value as the bond's yield would have to rise and its price fall to compensate for the increase in credit risk. The put option allows the investor to receive the full face value of the bond, plus accrued interest. Since this protection is valuable, investors must pay extra for it. Issuers can lower their debt costs by attaching put provisions to their bond issues.

176. Put Bond

A bond that the holder may choose either to exchange for par value at some date or to extend for a given number of years.

177. Put Option

A put option gives the holder the right to sell a certain number of shares of common stock at a price on or before the expiration date of the option. In purchasing a put, the owner of the shares has bought the right to sell those shares by the expiration date at the exercise price. As with calls, one can create, or write, a put, accepting the obligation to buy shares.

178. Put Provision

Gives holder of a floating-rate bond the right to redeem his or her note at par on the coupon payment date.

179. Put-Call-Parity

A relationship stating that the difference between the premiums of a call and a put with the same strike price and time to expiration equals the difference between the present value of the forward price and the present value of the strike price.

$$C + Xe^{-rT} = P + S$$

where C is defined as the call price per share; X is the strike price; P is the Put price per share; S is the stock price per share.

180. Put-Call Parity for American Options

Put-call parity applies only to European options, which can only be exercised on the expiration date, and not American options, which can be exercised before. It is complex to price American options since they can be exercised at any time point up to expiry date. The time the holder chooses to exercise the options depends on the spot price of the underlying asset. In this sense the exercising time is a random variable itself. It is obvious that the Black-Scholes differential equations still hold if the options are not exercised. However, the boundary conditions are so complicated that an analytical solution is not possible.

181. Put-Call Parity for Futures

In an efficient market, a portfolio that holds both a long call option and a short put option for the same asset, strike price and expiration date should generate the same return as a portfolio that holds an equivalent long position futures contract. This parity implies that the call option on futures is equal to the put option on futures plus the present value of futures minus the present value of the exercise price.

182. Puttable Bond

A bond where the holder has the right to sell it back to the issuer at certain predetermined times for a predetermined price.

183. Puttable Swap

A swap where one side has the right to terminate the swap early. On the other hand, a swap which one party has the option to extend the life of swap beyond the specified period is called extendable swap.

1.17 Q

1. Q Ratio or Tobin's Q Ratio

Market value of firm's assets divided by replacement value of firm's assets. It can be approximated by market/book ratio.

2. Quadruple Witching

Quadruple witching is the term used in the United States to describe the third Friday of March, June, September, and December. These are the days when index

futures, stock futures, index futures options, and stock options all reach their final trading day. Although the final trading day for most stock options is the third Friday of every month, it is the third Friday of the last month of each quarter for the other three instruments. Therefore, the third Friday in each of the months mentioned sees the last trading day for four derivative instruments – which is why the word quadruple is used.

3. Quality Financial Statements

Analysts sometimes speak of the *quality* of a firm's earnings, or the quality of its balance sheet. In general, quality financial statements are those that accurately reflect reality; they lack accounting tricks and one-time changes designed to make the firm appear stronger than it really is. Financial statements reflect reality when accounting income is a good approximation to economic income.

The Balance Sheet

A quality balance sheet typically shows conservative use of debt or leverage, which keeps the potential of financial distress due to debt service quite low. Limited use of debt also implies the firm has unused borrowing capacity; should an attractive investment opportunity arise, it can draw upon that unused capacity to invest wisely for the shareholders' benefit.

A quality balance sheet shows assets whose market values exceed their book values. In general, inflation and historical cost accounting should keep book values below market values. Beyond these accounting effects, a capable management team and the existence of intangible assets, such as goodwill, trademarks, or patents, will make the market values of firm's assets exceed their book values. Situations that might reduce assets' market values below their book values include: use of outdated, technologically inferior assets; unwanted out-of-fashion inventory; and the presence of nonperforming assets on the firm's books (as when a bank writes off a nonperforming loan).

The presence of off-balance sheet liabilities also harms the quality of a balance sheet by hiding economically important information. Such liabilities may include joint ventures and loan commitments or guarantees to subsidiaries.

The Income Statement

High quality earnings are *recurring* earnings that arise from sales to the firm's regular stream of customers. One-time and nonrecurring effects, such as accounting changes, mergers, and asset sales, should be ignored when examining earnings. Also, costs must not appear artificially low as a result of unusual and short-lived input price reductions. Unexpected exchange rate fluctuations that work in the firm's favor to raise revenues or reduce costs also should be viewed as nonrecurring.

Quality earnings are revealed by conservative accounting principles that do not overstate revenues or understate costs. The quality of the income statement rises as its statement of earnings more closely approximates cash. Suppose that a firm sells furniture on credit, allowing customers to make monthly payments. A high-quality income statement should recognize this revenue using the installment principle

(i.e., the statement of sales revenue should reflect only the cash collected from sales each month during the year). A low-quality sheet would recognize 100% of the revenue from a sale at the time of sale, even though payments may stretch well into the following year. The footnotes to the income statement would tell the analyst which method was used.

4. Quality Risk

The exact standard or grade of the commodity required by the hedger is not covered by the futures contract. Therefore, the price movement of commodity grade A may be different from the price movement of commodity grade B, which will cause the basis to change and prevent the hedger from forming a perfect hedge.

5. Quality Spread

The difference in market yields between yield on risky securities and matched maturity/duration Treasury securities.

6. Quantile

The percentage of data points below a given value. The q th quantile of the distribution F is the smallest value x such that $F(x) (\geq) q$.

7. Quantity Risk

The exact amount of the commodity needed by the hedger is not available by a single futures contract or any integer multiple thereof. Hence, the amount of the commodity is not hedged exactly; this prevents the hedger from forming a perfect hedge, and the underhedged or overhedged amount is subject to risk.

8. Quick Assets

Current assets minus inventories.

9. Quanto (Cross Currency Derivative)

A derivative where the payoff is defined in terms of value of variables associated with one currency but is paid in another currency. Therefore, this kind of derivative is a cross-currency derivative.

10. Quasi-Arbitrage

The replacement of one asset or position with another that has equivalent risk and a higher expected rate of return. This is an implicit instead of an explicit arbitrage.

11. Quasi-Random Sequence

A quasi-random sequence (also called a low discrepancy sequence) is a sequence of representative samples from a probability distribution. Descriptions of the use of quasi-random sequences can be found in Brotherton-Ratcliffe (1994, pp. 53–58).

Quasi-random sampling is similar to stratified sampling. The objective is to sample representative values for the underlying variables. In stratified sampling it is assumed that we know in advance how many samples will be taken. A quasi-random sampling scheme is more flexible. The samples are taken in such a way that we are always “filling in” gaps between existing samples. At each stage of the simulation the points sampled are roughly evenly spaced throughout the probability space. [See **Stratified sampling**]

12. Quick (Acid-Test) Ratio

A measure of liquidity similar to the current ratio except for exclusion of inventories in the numerator. The formula is:

$$\frac{\text{Cash} + \text{Receivables} + \text{Marketable Securities}}{\text{Current Liabilities}}$$

The quick ratio is a better measure of liquidity than the current ratio for firms whose inventory is not readily convertible into cash. [See **Liquidity ratios**]

1.18 R

1. R-Squared (R^2)

Square of the correlation coefficient proportion of the variability explained by the linear model. R^2 for regression for estimating beta is $\left(R^2 = \beta_i^2 \frac{\text{var}(R_m)}{\text{var}(R_{it})}\right)$. Under this case, R^2 represents the ratio between systematic risk and total risk. [See **Beta and market model**]

2. Rainbow Option

An option that has a payoff based on the maximum or minimum of two (or more) risky assets and cash. For example, the payoff to a rainbow call is $\max(S_t, Q_t, K)$, where S_t and Q_t are risky asset prices. This kind of option is often called two-color rainbow option, because the maximum and the minimum prices of two assets look very much like the shape of rainbow in a two-dimensional diagram, with two asset prices as the two axes.

3. Random Equation

The Ito equation is a random equation. The domain of the equation is $[0, \infty) * \Omega$ with the first argument t denoting time and taking values continuously in the interval $[0, \infty)$, and the second argument w denoting a random element taking values from a random set Ω . The range of the equation is the real numbers or real vectors. For simplicity only the real numbers, denoted by R , are considered as the range of equation.

$$dS(t.w) = \mu[t, S(t, w)]dt + \sigma[t, S(t, w)]dZ(t, w)$$

4. Random Walk

Theories that stock price changes from day to day are at random; the changes are independent of each other and have the same probability distribution. Mathematically, it is a stochastic process, $X(t)$, in which increments, $e(t)$, are independent and identically distributed:

$$X(t) = X(t - h) + e(t) \quad (24)$$

Under the weak-form efficient market, the relationship between stock prices per share is period $t(P_t)$ and that in period $t - 1$ (P_{t-1}) can be defined as:

$$P_t = P_{t-1} + \text{expected return} + \text{random error}(e_t) \quad (25)$$

If the stock prices follow Eq. 25, they are said to follow random walk.

5. Range-Forward Contract

Range-forward contract consists of a long forward contract combined with a long position in a put and a short position in a call. The strike prices are chosen so that the initial value of the call equals the initial value of the put. Since the value of the forward contract is zero initially, the value of the whole package is also zero. A range forward contract has a similar type of payoff pattern to a bull spread. [See **Bull spread**]

6. Rank Order

A quality of data often found across credit rating categories where values consistently progress in one direction – never reversing direction. Mathematicians term this property of data, *monotonicity*.

7. Rate Anticipation Swap

A switch made in response to forecasts of interest rates.

8. Rate Cap

A maximum interest rate inserted into a loan agreement which limits how far the loan interest rate can raise over the term of the loan.

9. Rate Collar

A loan agreement containing both a minimum and a maximum loan rate level to protect both borrower and lender against excessive interest rate risk.

10. Rate of Interest

The price of acquiring credit usually expressed as a ratio of the cost of securing credit to the total amount of credit obtained.

11. Rate Sensitive

Classification of assets and liabilities that can be repriced within a specific time frame, either because they mature or carry floating or variable rates.

12. Rate-Hedging Methods

Techniques for insulating an investor's holdings of financial assets from loss due to changes in market rates of interest.

13. Rating

System of assigning letters to security issues indicating the perceived default risks associated with that class of issues. Rating agencies include Standard & Poor and Moody etc. [See **Bond rating**]

14. Ratings Transitions

A change in the credit rating of a bond from one value to another. For example, a rating downgrade is from AAA to AA by Standard and Poor.

15. Ratio Analysis

Ratio analysis is another means by which to gain insight regarding a firm's strengths and weaknesses. Ratios are constructed by dividing various financial statement numbers into one another. The ratios then can be examined to determine trends and reasons for changes in the financial statement quantities. Ratios are valuable tools, as they standardize balance sheet and income statement numbers; thus, differences in firm size will not affect the analysis.

Three basic categories of ratio analysis typically are used, **time series analysis**, **cross-sectional analysis**, and **benchmark analysis**. [See also **Time series analysis**, **Cross-sectional analysis**, and **Benchmark analysis**]

There are also many categories of financial ratios. The following list represents the most basic categories:

1. Liquidity ratios
2. Asset management ratios
3. Capital structure ratios
4. Profitability ratios
5. Market value ratios

16. Ratio Spread

Buying m calls at one strike price and selling n calls at a different strike price, with all options having the same time to maturity and same underlying asset.

17. Rational Expectation Theory of Interest Rates

An explanation of the level of and changes in interest rates based on changes in investor expectations regarding future security prices and returns.

18. Real Assets, Financial Assets

Real assets are land, buildings, and equipment that are used to produce goods and services. *Financial assets* are claims such as securities to the income generated by real assets.

19. Real Cash Flow

A future cash flow of capital budgeting decision is expressed in real terms if the current, or date 0, purchasing power of the cash flow is given. In other words, it is the nominal cash flow divided by $1 + \text{inflation rate}$.

20. Real Earnings Management

Real earnings management refers to company management in real terms. Such as, cost of goods sold and other production expenses to manipulate the company's earnings results.

21. Real Estate Investment Trust (REITs)

There are two alternative ways to define REITs:

- (a) Tax-exempt corporations that receive at least three quarters of their gross income from real estate transactions and devote a high percentage of their assets to real property loans.
- (b) Real estate investment trust, which is similar to a closed-end mutual fund. REIT's invest in real estate or loans secured by real estate and issue shares in such investments. [See **Closed-end fund**]

22. Real Interest Rate

Interest rate expressed in terms of real goods; that is, the nominal interest rate minus the expected inflation rate. [See **Nominal risk-free interest**]

23. Real Investment

The net change in real assets held by a sector or unit in the economy over a specified time period.

24. Real Option

Option involving real (as opposed to financial) assets. Real assets include land, plant, and machinery. Similar to options on financial securities, real options involve discretionary decisions or rights, with no obligation, to acquire or exchange an asset for a specified alternative price. The ability to value real options (e.g., to defer, expand, contract, abandon, switch use, or otherwise alter a capital investment) has brought a revolution to modern corporate resource allocation.

25. Real Risk-Free Rate of Interest

The real risk-free interest rate is the return investors require on a zero-risk instrument with no inflation. Since no such security or economic environment exists, the real risk-free rate is admittedly a theoretical concept. It forms the basis for all expected returns and observed interest rates in the economy. Although it cannot be observed directly, it can be estimated. Studies indicate that, over time, the real risk-free interest rate in a country is approximately equal to the economy's long-run growth rate. But short-term influences can lead to increase or reductions in the real risk-free rate.

For example, short-term increases in growth above long-term trends (e.g., the business cycle) can cause an economy to have a larger demand for capital than a low-growth or recessionary economy. Larger government budget deficits are an additional source of demand for capital; all else being equal, they lead to higher real risk-free interest rates.

Supply forces can affect the real risk-free rate as well. Changes in national savings affect the pool of funds available for investment. Actions by the Fed can affect the short-term supply of capital and real interest rates. As people typically spend more

than they earn when they are young and then earn more than they spend as they grow older, the graying of the baby boomers in the United States may boost the supply of capital through a positive influence on personal savings. Legislation offering tax shields or other inducements to save, such as individual retirement accounts (IRAs) and tax-deferred annuities, also increase savings and the supply of capital.

26. Realized Compound Yield

A measure of total return calculated by comparing total future dollars equal to coupon interest or dividends plus reinvestment income and maturity or sale value of the underlying asset, with the initial purchase price, over the appropriate number of compounding periods.

27. Rebalancing

The process of adjusting a trading position periodically. Usually the purpose is to maintain delta neutrality. In the case of options and other more complicated derivatives, the hedge that is set up is only instantaneously riskless. To remain riskless it must be rebalanced continuously. [See **Delta hedging**]

28. Rebate

The return of a portion of unearned interest to a borrower.

29. Rebate Option

A claim that pays \$1 at the time the price of the underlying asset reaches a barrier. [See **Deferred rebate option**]

30. Receivables

Account receivable which are non-interest bearing short term extensions of credit to customer in the normal course of business. This kind of “trade credit” might be at risk to the extent that the customer may not pay its obligation in full.

31. Receivable Balance Pattern

The receivable balance pattern, also known as the **payments pattern approach**, provides a way of monitoring accounts receivable. This technique examines the percentage of credit sales for a given time period (usually 1 month) that are still outstanding at the end of each subsequent time period.

This approach is not affected by changes in sales levels, as the average collection period (ACP) and aging schedule are, so the receivable balance pattern does not give misleading signals. In addition, this approach can develop predictions of receivable balances and collections as part of a cash flow forecast.

32. Receivables Turnover Ratio

Total operating revenues divided by average receivables. Used to measure how effectively a firm is managing its accounts receivable. [See **Asset management ratios**]

33. Receiver Swaption

A swaption giving the holder the right to receive the fixed rate in a swap. Thus, the holder of a receiver swaption would exercise when the fixed rate is below the strike price.

34. Recombining Tree

A binomial tree describing asset price moves in which an up move followed by a down move generates the same stock price as a down move followed by an up move. It is also called a lattice. Binomial option pricing model is based upon a lattice.

35. Record Date

The shareholders whose names appear on the corporation's list of shareholders on the record date are entitled to receive the dividend even if they sell their stock before the payment date. [See also **Dividend declaration date**]

36. Recourse

Legal right to enforce a claim against another party.

37. Recoveries

The dollar amount of loans that were previous charge off but now collected.

38. Recovery Value

The percentage of par value received by either a bondholder or a lender in a bankruptcy.

39. Recovery Rate

It can be defined as $1 - \text{loss given default (LGD)}$. Recovery rates for individual obligors differ by issuer and industry classification. Rating agencies such as Moody's publish data on the average prices of all defaulted bonds, and generally analysts will construct a database of recovery rates by industry and credit rating for use in modeling the expected recovery rates of assets in the collateral pool.

40. Red Herring

A red herring is basically a preliminary **prospectus**. [See also **Prospectus**] The nickname arises from the disclaimer, printed in red on the cover of the prospectus, that the SEC has not yet approved the securities for sale.

41. Redlining

A practice whereby lenders deny loans to residents living in predetermined geographic areas. For example, many lenders were found to be making mortgage loans much more readily available in white neighbors than in those with higher proportions of nonwhites. Such a practice is illegal.

42. Reference Price

A market price or rate used to determine the payoff on a derivatives contract.

43. Refunding

The process of replacing outstanding bonds, typically to issue new securities at a lower interest rate than those replaced.

44. Registered Bond

A bond whose issuer records ownership and interest payments. Differs from a bearer bond, which is traded without record of ownership and whose possession is its only evidence of ownership. [See **Bearer bond**]

45. Registered Trader

A member of the exchange who executes frequent trades for his or her own account.

46. Registration Statement

The registration that discloses all the pertinent information concerning the corporation that wants to make the offering. The statement is filed with the Securities and Exchange Commission.

47. Regression Equation

An equation that describes the average relationship between a dependent variable and a set of explanatory variables. [See **Scatter diagram** and **Market model**]

48. Regular Cash Dividend

Cash payment by firm to its shareholders, usually four times a year.

49. Regulation A

The securities regulation that exempts small public offerings (those valued as less than \$1.5 million) from most registration requirements.

50. Regulation Q Ceilings

Maximum interest rates on certain types of deposits that are set under federal law and administered by a regulatory agency such as the Federal Reserve Board.

51. Reinvestment Rate Risk

The return that an investor receives from a bond investment equals the bond's yield to maturity or effective annual rate only if the coupon payments can be reinvested at a rate equal to the bond's yield to maturity. Since the form of the interest factor in the bond price equation, $(1 + r)^m$ assumes that all the cash flows are reinvested at the periodic rate r ; should future coupons be reinvested at a lower rate, the investor's actual yield will be less than the bond's yield to maturity. Therefore, reinvestment rate risk occurs when fluctuating interest rates cause coupon payments to be reinvested at different interest rates. Another illustration of reinvestment rate risk occurs when maturing bank CDs are rolled over into new CDs. The risk benefits the investor when the new CD rate is higher than the maturing CD rate; it works against the investor when the new CD rate is lower.

A zero coupon bond, a bond which pays no explicit interest, eliminates reinvestment risk. This is the primary reason for the popularity of zero coupon bonds in the investors.

52. Reinvestment Risk

The risk that future cash flows may be reinvested at rates below those expected or available at present. [See **Reinvestment rate risk**]

53. Relative Price Risk

Relative price risk is one type of the exchange rate risk. It is due to changes in supply-and-demand conditions in various countries.

54. Relative Purchasing Power Parity

A more useful offshoot of **absolute purchasing power parity** is relative purchasing power parity. [See also **Absolute purchasing power parity**] Relative purchasing

power parity claims that the exchange rates between countries will adjust over time to reflect their relative inflation rates. If h_{FC} and h_{US} are the inflation rates in a foreign country and the United States, respectively, relative purchasing power parity claims that the expected change in the spot rate between the currencies (ΔER) is given by:

$$\Delta ER = \frac{E(S_1)}{S_0} - 1 = \frac{h_{FC} - h_{US}}{1 + h_{US}}$$

which is equivalent to:

$$1 + \Delta ER = \frac{E(S_1)}{S_0} = \frac{1 + h_{FC}}{1 + h_{US}}$$

where S_0 and $E(S_1)$ are the current spot exchange rate and the expected spot rate 1 year in the future, respectively.

In sum, relative purchasing power parity (RPPP) is the idea that the rate of change in the price level of commodities in one country relative to the price level in another determines the rate of the exchange rate between the two countries' currencies. [See **International Fisher Effect**]

55. Remainder Man

One who receives the principal of a trust when it is dissolved.

56. REMIC

A real estate mortgage investment conduit [REMIC] issuing securities collateralized by mortgages (CMOs) and passing on principal and interest payments to investors. Like CMOs, REMIC securities represent claims on the underlying cash flows that are prioritized by multiple classes or branches.

57. Reorganization

Financial restructuring of a failed firm. Both the firm's asset structure and its financial structure are changed to reflect their true value, and claims are settled. Current law allows the bankrupt firm to be reorganized under ► [Chap. 12, "Working Capital and Cash Flow"](#). The objective of reorganization is to keep the firm alive while settling creditor's claims and attracting new capital into the firm. [See ► [Chapter 12, "Working Capital and Cash Flow"](#)]

58. Replacement Cost

Cost to replace a firm's assets. "Reproduction" cost.

59. Replacement Value

Current cost of replacing the firm's assets.

60. Replacement-Chain Problem

Idea that future replacement decisions must be taken into account in selecting among projects.

61. Repo

Repurchasing agreement. A procedure for borrowing money by selling securities to a counterparty and agreeing to buy them back later at a slightly higher price. [See **Repurchase agreement**]

62. Repo Rate

The annualized percentage difference between the original sale price and final repurchase price in a repurchase agreement.

63. Representative Offices

Facilities established in distant markets by a bank in order to sell the bank's services and assist its clients; these offices usually cannot accept deposits or make loans.

64. Repricing

The replacement of an out-of-the-money compensation option with an at-the-money compensation option. This kind of reducing the exercise price of compensation option in response to a decline in stock price is called option repricing.

65. Repurchase Agreements

Repurchase agreements (repos) are not actual securities in themselves, but rather contracts to immediately acquire available funds by selling securities, together with a simultaneous agreement to repurchase those securities at a later date. Most repos are outstanding for only one business day, and nearly all involve Treasury or government agency securities.

For example, suppose a company has \$1 million in excess cash available for 2 days. Instead of buying T-bills and then selling them 2 days later, the company could create a repurchase agreement with a bank. The company would agree to purchase \$1 million worth of T-bills and then sell them back to the bank after 2 days for the original \$1 million plus 2 days of interest. No actual transfer of physical securities is made; rather, the entire transaction consists of bookkeeping entries on the two parties' accounts.

Repos offer two distinct advantages for investing short-term surplus cash. First, their maturities can be tailored to suit the exact times that the parties have funds available, from overnight to 30 days or more. Second, because repos state the selling

price of the securities in the initial agreement between buyer and seller, they eliminate interest rate risk. The yields on repos are similar to, but slightly lower than, those of T-bills.

66. Repurchase of Stock

Device to pay cash to firm's shareholders that provides more preferable tax treatment for shareholders than dividends. Treasury stock is the name given to previously issued stock that has been repurchased by the firm. [See **Stock repurchase**]

67. Required Reserves

Holdings of cash and funds on deposit with the Federal Reserve banks by depository institutions that are required by law to backstop the public's deposits held by these same institutions.

68. Reserve Cash

A company's cash needs fall into three categories: (1) cash for day-to-day transactions, (2) reserve cash to meet contingencies, and (3) cash for compensating balance requirements. To estimate reserve cash requirements, the cash flow manager can tabulate the daily or weekly changes in the cash amount. These changes will range from some very large changes to small fluctuations. Because the major cash flow problem is running short of cash, the financial manager is especially interested in large decreases and thus might select a reserve balance that would meet all but the largest historic cash decreases.

69. Reserve for Bank Debts

Amount appearing on a bank's balance sheet that represents the estimated value of uncollected loans.

70. Reserve Requirements

In regulating the banking industry, the Fed set reserve requirements to specify the portion of a bank's total deposits that it must hold as reserves. The Fed hesitates to change the reserve requirement due to the money multiplier effect; small changes in the reserve ratio can have a very large impact on money supply. This makes the tool too coarse for the subtle work of adjusting the economy.

71. Reserve Requirement Ratios

Percentages applied to transactions account and time deposits to determine the dollar amount of required reserve assets.

72. Reserve Target

The minimum daily reserve ratio of deposit institution (DI) required by Fed. In general, the DI can either undershoot or overshoot this ratio.

73. Reserves

Qualifying assets to meet reserve requirements, including vault cash and deposit balances help at Federal Reserve banks.

74. Reset Date

The date in a swap or cap or floor when the floating rate for the next period is set.

75. Residential Mortgages

Loans secured by single-family homes and other dwellings.

76. Residential Mortgage Credit

Loans provided to support the purchase of new or existing single-family homes, duplexes, and other permanent dwellings.

77. Residual Claim

Refers to the fact that shareholders are at the bottom of the list of claimants to assets of a corporation in the event of a failure or bankruptcy.

78. Residual Dividend Approach

An approach that suggests that a firm pay dividends if and only if acceptable investment opportunities for those funds are currently unavailable. [See **Residual theory of dividends**]

79. Residual Theory

The most easily understood theory of dividend payment determination is called the residual theory. As the name implies, this theory holds that firms pay dividends out of earnings that remain after it meets its financing needs. These are funds for which the firm has no immediate use. The procedure for a residual dividend policy follows several steps:

1. Determine the firm's optimal capital budget
2. Determine the amount of equity needed to finance that budget

3. To the extent possible, use the firm's retained earnings to supply the needed equity
4. Distribute any leftover earnings as dividends

The basic assumption of residual dividend theory is that shareholders want the firm to retain earnings if reinvesting them can generate higher rates of return than the shareholders could obtain by reinvesting their dividends. For example, if a corporation can invest retained earnings in a new venture that generates an 18% rate of return, whereas investors can obtain a return of only 10% by reinvesting their dividends, then stockholders would benefit more from the firm reinvesting its profits.

Whether firms actually practice the residual theory is a matter of question. Such a theory would imply erratic dividend payments, especially for fast-growth companies. Firms do seem to try to stabilize their dividend-payout rates, so analysts do not place much faith in the residual theory. However, two alternative theories for the dividend behavior of firms have found considerable empirical support.

80. Residual Value

Usually refers to the value of a lessor's property at the time the lease expires.

81. Residuals

Parts of stock returns not explained by the explanatory variable (the market-index return). They measure the impact of firm-specific events during a particular period. [See also **Market model**]

82. Resistance Level

A price level above which it is supposedly difficult for a stock or stock index to rise.

83. Resolution Trust Corporation (RTC)

A government agency (1989–1996) that assisted in the management and savings and loans deemed to be insolvent during the Thrift Crisis. At the time of its dissolution in 1995, RTC had resolved or closed more than 700 saving institution.

84. Respondent Bank

Bank that purchases services from a correspondent bank.

85. Restrictive Covenants

Provisions that place constraints on the operations of borrowers, such as restrictions on working capital, fixed assets, future borrowing, and payment of dividend.

86. Retained Earnings

Earnings not paid out as dividends. It is one of the items of equity statement. This term also appears in the balance sheet.

87. Retention Rate

The retention rate represents the proportion of every \$1 of earnings per share that is retained by the firm; in other words, it is equal to one minus the **dividend payout ratio**.

88. Retention Ratio

Retained earnings divided by net income. It is equal to one minus payout ratio.

89. Retractable Bonds

[See **Puttable bonds**]

90. Return

Profit on capital investments or securities.

91. Return Items

Checks that have not been honored by the drawee bank and have been returned to the check writer.

92. Return on Assets [ROA]

Income divided by average total assets. [See **Profitability ratios**]

93. Return on Equity (ROE)

Net income after interest and taxes divided by average common stockholders' equity. [See **Profitability ratios**]

94. Return on Sales (ROS), or Profit Margin

The ratio of operating profits per dollar of sales (EBIT divided by sales). [See **Profitability ratios**]

95. Revenue Bond

Most of municipal bonds are revenue bonds. They will be repaid out of proceeds from the specific revenue-generating project that they were sold finance, such as toll roads.

96. Revenue Surprises

Revenue is the total amount of money a company generates in the course of its normal business operations. Generally, analysts and investors carefully assess the company's revenues from different periods to identify their growth trends. In some cases, the reliability of revenue can be questionable as the metric is prone to potential manipulation. For example, the management of a company can artificially inflate revenues by applying aggressive revenue recognition principles.

97. Revenue-Anticipation Notes (RANs)

Short-term debt obligations issued by state and local units of government in lieu of expected future governmental revenues in order to meet near-time cash demands.

98. Reverse Cash-and-Carry

The simultaneous short-sale and forward purchase of an asset or commodity.

99. Reverse Conversion

A short position in an asset coupled with a purchased call and written put, both with the same strike price and time to expiration. The position is a synthetic short T-bill position.

100. Reverse Mortgage

A mortgage in which the owner of the property can borrow against existing equity in the property.

101. Reverse Purchase Agreement

The purchase of a security coupled with an agreement to sell it at a later date. The opposite of a repurchase agreement.

102. Reverse Repo

A contract in which a lender provides funds to a borrower for which collateral is provided in the event of nonpayment. Every RP transaction involves both a regular RP and a reverse RP depending on whether it is viewed from the lender's or borrower's prospective. Most RPs use for Treasury or US agency securities as collateral.

103. Reverse Repurchase Agreement

Securities purchased under an agreement to resell them at a later date.

104. Reverse Split

The procedure whereby the number of outstanding stock shares is reduced; for example, two outstanding shares are combined to create one. [See **Stock split**]

105. Reverse Stock Split

A reverse stock split, as its names implies, is a reduction in the number of shares outstanding, with each share increasing in value to keep the total value of the firm unchanged. As with a **stock split**, theory gives no reason to expect any change in the underlying value of the company that engages in a reverse split. [See also **Stock split**] In fact, many investors regard a reverse split as an admission by management that the company faces financial difficulties. This belief is based primarily on the argument that the market price per share is too low to attract serious investors.

106. Reversible Swap

Allows counterparty to change status from floating-rate payer to fixed-rate payer and vice versa.

107. Reversing Trade

Entering the opposite side of a currently held futures position to close out the position.

108. Reversion Level

The level to which the value of a market variable (e.g., an interest rate) tends to revert.

109. Revolving Commitment (revolver)

A generic term referring to some facility which a client can use – or refrain from using – without canceling the facility. In other words, it guarantees that funds can be borrowed, repaid and borrowed again over an extended period, perhaps as long as 3 years. [See **Revolving credit agreement**]

110. Revolving Credit Agreement

Banks usually grant lines of credit for specific lengths of time, usually 1 year or less. The parties may, of course, renegotiate the loan to provide the funds for a longer time, if needed. Still, the bank usually expects the borrower to clean up the loan – that is, to reduce its debt to the bank to zero – at least once during the year.

A borrower that has a recurring need for funds may instead arrange a revolving credit agreement. This type of loan resembles the line of credit, in that the parties

agree to a maximum credit level, and the borrower may draw funds up to that limit. The revolving credit agreement, however, meets the borrower's need to borrow the funds, pay off the loan, and then borrow again, time after time. Such a situation may supply funds for a borrowing company that produces a small number of large, high-value products, such as ships or steam turbines; the firm must borrow to finance the construction of each product until it eventually collects the proceeds of the sale. Moreover, a revolving credit agreement is more likely to be guaranteed by the bank than a line of credit.

Because the bank must commit to the agreement for a much longer time than a conventional line of credit would demand, the negotiation process for a revolving credit agreement tends to be more formal. The bank may specify that the borrower must maintain its working capital above a specified level, forbid any factoring of accounts receivable without the bank's permission, or stipulate that any further borrowing must be subordinated to the revolving credit debt. Commitment fees also are common for large revolving credit agreements. Most banks offer the borrower a choice between a **committed line of credit** and an **uncommitted line of credit**. [See also **Committed line of credit** and **Uncommitted line of credit**]

111. Revolving Loan

A credit line on which a borrower can both draw and repay many times over the life of the loan contract. [See also **Revolving credit agreement**]

112. Reward-to-Volatility Ratio

Ratio of excess return to portfolio standard deviation. [See **Sharpe ratio**]

113. Rho

The change in value of a derivative due to a change in the interest rate. Based upon the call option formula defined in option pricing model (see section "O"). The mathematical result can be defined as:

$$\frac{\partial C}{\partial r} = TXe^{-rT}N(d_2) > 0$$

114. Riding the Yield Curve

It is one of the bond portfolio management strategies designed to increase income is called **riding the yield curve**. To be successful, managers using this approach must be willing to make several rather strong assumptions about the future course of interest rates. To illustrate, suppose that a manager would ordinarily hold 1-year Treasury securities as part of an institution's secondary reserves but now sees an upward-sloping yield curve. The yield on 1-year Treasuries is 3.5% and the yield on 2-year Treasuries is 4.0%. *If the manager assumes that the shape and level of the*

yield curve will remain the same, the price of 2-year Treasuries must rise so that their yield next year (when they will be 1-year Treasuries) will be 3.5%. A manager willing to ride the yield curve would hold 2-year Treasuries this year, then sell them at the end of the year after their price rose to provide additional income on the portfolio. Assuming that the level and shape of the yield curve has, in fact, remained unchanged, the manager would then reinvest the proceeds in 2-year Treasuries and begin the ride again.

In sum, riding the yield curve is an investment strategy the investor buys a security that matures after the investor's assumed holding period. The investor plans to sell the security at the end of the holding period and earn an above-average return because interest rates are expected to remain stable or fall.

115. Rights Issue

An issue to existing shareholders of a security giving them the right to buy new shares at a certain price.

116. Rights Offering

A rights offering allows the firm's current shareholders to purchase additional shares in proportion to their current ownership. This way, the original shareholders maintain control of the firm while raising the needed equity capital among themselves. For example, suppose four shareholders each own 25% of a firm whose equity has a value of \$4 million. If the firm needs an additional \$1 million to finance a plant expansion, it can make a rights offering, allowing each shareholder to invest \$250,000. This way, each shareholder can retain 25% ownership in the expanded firm.

If a shareholder cannot or declines to invest the full \$250,000, the remaining shareholders can invest the difference. Proportionate ownership would change to reflect the overall fraction invested by each shareholder.

The advantage of the rights offering is that the current set of shareholders can maintain control of the firm by each contributing additional funds to meet the firm's needs. This condition creates a practical difficulty, though. As a group, the shareholders may not be able to raise the needed funds, leading to a failure of the rights offering. The firm then must arrange financing from other sources of private equity.

117. Risk Arbitrage

Speculation on perceived mispriced securities, usually in connection with merger and acquisition targets.

118. Risk Assessment

Risk assessment serves many purposes for an organization, including reducing operational risks, improving safety performance, and achieving objectives. While many individuals are involved in the process and many factors come into play,

performing an effective risk assessment comes down to three core elements: risk identification, risk analysis, and risk evaluation.

119. Risk Aversion

A rational investor prefers returns and dislikes risk. Therefore, this kind of investor is a risk-averse investor.

120. Risk Classification

Certain types of projects are inherently more or less risky than other. The firm can use past experience and information from audits of earlier projects to create risk classes or categories for different types of capital budgeting projects. The findings from break-even, scenario, sensitivity, or simulation analysis also can be used to determine risk categories for projects. Each risk category can be given a generic description to indicate the types of projects it should include and a risk-adjusted discount rate or project cost of capital to assign those projects. An example is shown in Table 8.

Subjectivity enters this process as management must decide the number of categories, the description of each risk category, and the required rate of return to assign each category.

Differences of opinion or internal firm politics may lead to controversy in classifying a project. Clearly defined category descriptions can minimize such problems.

The process of setting up the risk categories can be made less subjective if the firm audits ongoing and completed capital budgeting projects. Audits can provide fairly objective written records of the firm's experiences with different categories of projects. This paper analysis trail can be used to justify the classifications given to different kinds of projects, as well as the risk premiums assigned to different risk classes.

121. Risk Class

A partition of the universal set of risk measure so that projects that are in the same risk class can be comparable. M&M propositions have derived in terms of risk class assumption.

Table 8 Risk categories: XYZ corporation

Below-average risk	Replacement decisions that require no change, or only a minor change, in technology. No change in plant layout required. Discount rate = cost of capital - 2%
Average risk	Replacement decisions involving significant changes in technology or plant layout; all? cost-saving? decisions; expansions and improvements in the firm's main product lines. Discount rate = cost of capital
Above-average risk	Applied research and development; expansion of production or marketing efforts into developed economies in Europe and Asia. Discount rate = cost of capital + 2%
High risk	Expansion of production or marketing efforts into? less-developed? and emerging economies; introduction of products not related to any of the firm's current product lines. Discount rate = cost of capital + 5%

122. Risk Lover

[See **Risk-averse**]

123. Risk Management

The active use of derivatives and other techniques to alter risk and protect profitability.

124. Risk Management Tools

Financial devices (such as futures and options) which permit a borrower or lender of funds to protect against the risks of changing prices and interest rates.

125. Risk Neutral

A term describing an investor who is indifferent between receiving amount of x dollar and taking a risky bet with an expected value equal to x dollar. [See also **Risk-averse**]

126. Risk Premiums

The nominal risk-free return is the same for all investments throughout the market. However, there are as many different interest rates or expected returns as there are time horizons (ranging from 1 day to many years) and financial instruments (from passbook savings accounts to corporate stocks). The interest rates we observe in the economy differ from the nominal risk-free rate due to risk premiums. With the possible exception of Treasury bills, all investments are risky.

In sum, risk premium is the excess return on the risky asset that is the difference between expected return on risky assets and the return on risk-free assets.

127. Risk-Averse, Risk Neutral, Risk Lover

A *risk-averse* investor will consider risky portfolios only if they provide compensation for risk via a risk premium. A *risk-neutral* investor finds the level of risk irrelevant and considers only the expected return of risky prospects. A *risk lover* is willing to accept lower expected returns on prospects with higher amounts of risk.

128. Risk-Free Asset

An asset with a certain rate of return; often taken to be short-term T-bills.

129. Risk-Free Investment

A risk-free investment is one in which the investor is sure about the timing and amount of income streams arising from the investment. However, for most types of

investments, investors are uncertain about the timing and amount of income of their investments. The types of risks involved in investments can be quite broad, from the relatively riskless T-bill highly risky speculative stock.

130. Risk-Free Rate

The interest rate that can be earned with certainty. It is risk free in terms of default risk instead of inflation risk. [See **Nominal risk-free interest rate**]

131. Risk-Neutral Measure

The probability distribution for an asset transformed so that the expected return on the asset is the risk-free rate.

132. Risk-Neutral Probability

In the binomial model, the probability of an up move in the asset price such that the expected return on the asset is the risk-free rate.

133. Risk-Neutral Valuation

The valuation of an option or other derivative assuming the world is risk neutral. Risk-neutral valuation gives the correct price for a derivative in all worlds, not just in a risk-neutral world.

134. Risk-Return Trade-Off

If an investor is willing to take on risk, there is the reward of higher expected returns. Both security market line and capital market line are used to determine the risk return trade off. [See **Capital market line and Security market line**]

135. Riskless Portfolio

A combination of assets that earns the riskless rate of interest over the chosen investment horizon. The investment horizon is assumed to be one period; the duration of this period can be any length of time – an hour, day, week, and so on.

136. Risky Asset

An asset with an uncertain rate of return. For example, stocks are risky assets.

137. Risky Corporate Debt

Sometimes, options are used to value risky corporate debt. Because of the limited liability of stockholders, money borrowed by the firm is backed, at most, by the total value of the firm's assets. One way to view this agreement is to consider

that stockholders have sold the entire firm to debt holders but hold a call option with an exercise price equal to the face value of the debt. In this case, if the value of the firm exceeds the value of the debt, stockholders exercise the call option by paying off the bondholders. If the value of the firm is less than the value of the debt, shareholders do not exercise the call option, and all assets are distributed to the bondholders.

138. Roll Back

[See **Backwards induction**]

139. Rolling Options

Rolling is a technique that essentially involves closing one options position and opening another position on the same underlying security using options contracts with different terms. There are three different forms of rolling; rolling forward, rolling up, and rolling down. They each have their own uses. The main use of rolling is as a tool to manage positions, and it's typically to try and maximize profits.

140. Roth IRA

An individual retirement account introduced in 1998 that allows individuals whose wages and salaries are below a predetermined minimum to contribute after-tax income. The contributions grow on a tax-sheltered basis and thus are not taxed at withdrawal.

141. Round Lot

Common stock trading unit of 100 shares or multiples of 100 shares. When an individual wants to buy fewer than 100 shares, the order is turned over to an odd-lot dealer who will buy or sell from his own inventory.

142. Rule 415

In 1983, the SEC passed Rule 415, which allows firms to register security issues (both debt and equity) and then “put them on the shelf” for sale any time over the next 2 years. Once registered, the securities can be offered for sale by submitting a short statement to the SEC whenever the firm needs the funds or when market conditions are attractive. [See also **Shelf registration**]

143. Rule of 72

Divide 72 by the interest rate at which funds are invested. The value indicates how long it will take for the amount of funds invested to double in value.

144. Run on a Bank

Situation in which a large number of depositors lose confidence in the safety of their deposits and attempt to withdraw their funds.

1.19 S

1. S&P 500

[See **Standard and Poor Composite index**]

2. Safe Deposit Box

Privacy boxes for storage in a bank vault under lock and key.

3. Safe Harbor Lease

A lease to transfer tax benefits to ownership (depreciation and debt tax shield) from the lessee, if the lessee cannot use them, to a lessor that can.

4. SAIF

Savings Association Insurance Fund which insures deposits at savings and loans. This is one of the two insurance funds under FDIC. [See **BIF**]

5. Sale and Lease-Back Agreement

In a sale and lease-back agreement, the owner of an asset sells it and then leases it back. This method allows cash-strapped firms to sell valuable assets, but still retain their use.

6. Sales Forecast

A key input to the firm's financial planning process. External sales forecasts are based on historical experience, statistical analysis, and consideration of various macroeconomic factors; internal sales forecasts are obtained from internal sources.

7. Sales Terms and Collections

The fastest way to collect receivables is to ask for the money regularly. However, a company also can change its sales terms in an attempt to collect cash more quickly. Such a policy can take several forms, including (a) introduce discounts; (b) reduce credit terms; (c) emphasize cash sales; (d) accept credit cards; and (e) impose penalties for late payment.

8. Sales-Type Lease

An arrangement whereby a firm leases its own equipment, such as IBM leasing its own computers, thereby competing with an independent leasing company.

9. Sallie Mae

Student Loan Marketing Association which guarantees student loans. The asset structure of Sallie Mae heavily dominated by floating-rate standard loans and advances. Investor supplying funds to Sallie Mae preferred to lock in the high rate prevailing at that time. Therefore, Sallie Mae, pioneered swap program in the United States in 1982.

10. Sample-Function Analysis

Sample-function analysis regards a time series as an observed sample function representing a realization of an underlying stochastic process. Complicated parametric statistical-estimation procedures are used to determine the properties of time-series data.

11. Savings

The amount of funds left over out of current income after current consumption expenditures are made, or for a business firm, the current net earnings retained in the business instead of paid out to the owners.

12. Savings Banks

Depository institutions that are owned by their depositors and can be chartered by both the federal government and by some states.

13. Savings and Loan Associations

The predominant home mortgage lender in the United States, making predominantly local loans to finance the purchase of housing for individuals and families.

14. Scalper

Speculators are often distinguished by the time they hold their position. They can either be scalper or day trader. Scalper is a trader who holds positions for a very short period of time.

15. Scatter Diagram of a Regression

To estimate beta coefficient we regress rate of return of company i in period t (R_{it}) on market rate of return in period t (R_{mt}), then the regression model can be defined as:

$$R_{it} = \alpha_i + \beta_i R_{mt} + e_{it} \quad (26)$$

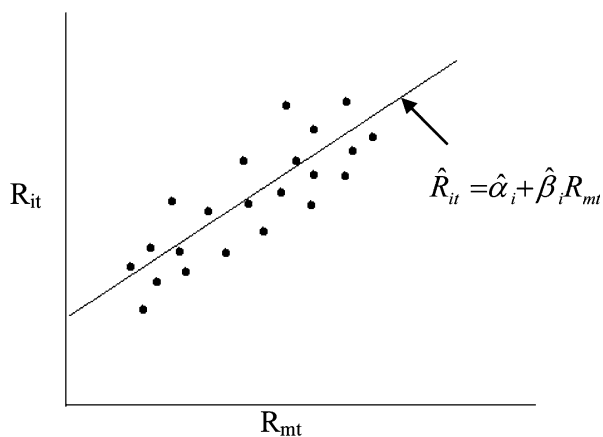
The estimated slope is the beta coefficient (systematic risk). In the regression analysis of Eq. 26 involving one independent variable (R_{mt}) and one dependent variable (R_{it}) the individual values of R_{it} and R_{mt} are plotted on a two-dimensional graph. In this two-dimensional graph, we can plot the different points in accordance with the pair-wise observations of R_{it} and R_{mt} to obtain the scatter diagram as presented in Fig. 1.

16. Scenario Analysis

Scenario analysis provides a means to evaluate the potential variability in a capital budgeting project's NPV. Scenario analysis computes several net present values for the project based on different scenarios. The initial capital budgeting analysis using the analyst's estimates of expected cash flows is called the *base-case scenario*. From this base case, typically at least two other scenarios are developed – a worst-case scenario and a best-case scenario – and NPVs are computed for each. The worst case NPV and the best case NPV give managers a likely range in which the project's NPV will fall. The purpose of scenario analysis is to examine the joint impact on NPV of simultaneous changes in many different factors.

The worst-case scenario should reflect project results under Murphy's law: If anything can go wrong, it will. Compared to the base case, the worst-case scenario will have lower sales volume, lower prices, higher costs, shorter product life, lower

Fig. 1 Scatter diagram of market model



salvage value, and so on. Rather than being an exercise in disaster forecasting, however, the worst-case scenario should reflect the circumstances that could reasonably be expected should the project be plagued with bad luck or bad analysis. Some of the firm's past failures can be used as models for developing the worst-case scenario. The resulting estimates of cash flows and NPV will reflect this pessimistic perspective.

The best-case scenario should illustrate how the project will turn out if everything works better than expected. The sales figures, prices, costs, and so on should incorporate the upper boundary of *reasonable* optimism. An unrealistic pie-in-the-sky scenario, however, will add little to the analysis. Spreadsheet packages can facilitate the analysis of different scenarios.

The analyst then presents decision makers with three sets of conditions, cash flows, and NPVs. The base-case represents an estimate of the most likely outcome; the worst-case and best-case scenarios illustrate the project's possible extremes. The NPVs of the worst- and best-case scenarios represent the potential range of the project's impact on shareholder wealth. If the worst-case scenario has a large, negative NPV, management may call for more analysis to see if the project can be modified to reduce its potential for severely decreasing shareholder wealth.

Another possibility is that management may decide that the project's best-case scenario is so attractive that it overcomes the project's downside risk. This may be the case for a project with encouraging engineering or market test results or a project that may propel the firm into a position of industry leadership.

17. Scoring System

Lenders use a credit scoring system, or a numerical system, to measure how likely it is that a borrower will make payments on the money he or she borrows. It is created by assigning scores to various attributes associated with the applicant's creditworthiness.

18. Season Dating

Arrangements for credit transactions include special terms for recurrent purchases and seasonal accounts. When a company routinely buys many items, the supplier commonly accounts for all sales during the month as occurring at the end of the month, eliminating the inconvenience of a separate payment for each delivery. These terms are stated as 5/10, EOM, net 60; that is, by paying the bill within 10 days of the end of the month (EOM), the customer will receive a 5% discount. Otherwise, full payment is due within 60 days of the invoice date. Manufacturers who produce seasonal goods often encourage customers to take early delivery by allowing them to delay payment until the normal ordering season, a type of credit arrangement known as season dating.

19. Seasonal Liquidity Needs

Cash flow needs that arise from predictable seasonal loan demands and deposit outflows.

20. Seasonal Swap

Notional principal may vary up or down over the life of the swap. Also known as roller coaster swap.

21. Seasonality

Patterns in the behavior of interest rates, with rate of increases during certain seasons of the year and decreases during other seasons, year after year.

22. Seasoned New Issue

A new issue of stock after the company's securities has previously been issued. A seasoned new issue of common stock can be made by using a cash offer or a rights offer. [See **Seasoned offering**]

23. Seasoned Offering

Not every public sale of stock by a corporation is an IPO. Corporate growth and/or high debt ratios require some public firms to return to the equity markets to raise funds. A new stock offering by an already public company is called a seasoned offering. Such offerings are easier for the investment bank and investors to price. Rather than estimating fair market value from accounting data, as in an IPO, investors can refer to daily listings of the market value of the company's shares.

A public company that needs an equity capital infusion faces several choices. It can increase its equity base by selling shares of common or preferred stock, and it can raise money in the US market or issue securities overseas.

Only the United States has a public financial market for preferred equity issues; other countries have not developed primary and secondary markets for preferred stock trading. As preferred equity increases a firm's equity base without diluting control, more and more foreign firms are issuing both fixed- and adjustable-rate preferred stock in the US markets.

Overseas tax and regulatory environments may make fund-raising cheaper for large US firms. Analysis of the Euroequity and Eurobond markets provides evidence of such cost advantages.

After a firm decides upon the form in which it will raise equity, it can market the new issue in several ways. It can sell the new shares to the public or to current shareholders or place them privately.

24. Seat

Another term for an exchange membership is a seat. A prospective member may buy a full seat, allowing him or her to trade any of the offered futures contracts. To encourage volume on newer or less liquid contracts, most exchanges usually also offer a partial seat. Permitting its owner to trade only a designated number of

contracts. Usually, to get onto an exchange to trade, an investor needs to buy or lease a seat from a current owner. The value of an exchange seat can vary substantially.

25. SEC

Securities and Exchange Commission, which was established under the authority of the Securities Act of 1933 and the Securities Exchange Act of 1934. The SEC regulates security firm such as Merrill Lynch and investment banking firm, such as Salomon Smith Barney and Goldman Sachs. [See **Securities and Exchange Commission**]

26. Secondary Capital

Limited life preferred stock, subordinated debt, and mandatory convertible securities not included as primary capital. Secondary capital is the Tier 2 capital, and primary capital is the Tier 1 capital.

27. Secondary Market

The secondary market is the resale market for securities. The issuing entity (corporation or government) is usually not involved in such transactions. In the secondary market, investors buy and sell securities among themselves. Without the liquidity that the secondary market provides, the primary market would be much less attractive because investors could not easily sell their securities.

The secondary market includes two forums for security trades: **organized exchanges** and an **over-the-counter market**. [See also **Organized exchanges** and **Over-the-counter market**] The New York Stock Exchange (NYSE) is an example of an organized exchange.

28. Secondary Securities

Financial claims, such as deposits, issued by a financial intermediary to raise loanable funds.

29. Second Mortgage

A claim against real property arising from a loan which is subordinated to (come after) the claim held by the holder of a first mortgage against the same property.

30. Second Mortgage Bond

A second mortgage bond has a secondary, or junior, claim on assets. [See also **First mortgage bond**]

31. Second-Pass Regression

A cross-sectional regression of portfolio returns on betas. The estimated slope is the measurement of the reward for bearing systematic risk during the period. This technique was proposed by Fama and Macbeth (1973, pp. 607–636).

32. Sector Influences

In addition to an overall market factor, various factors related to industry-type indexes are significant in explaining the returns generating process for a particular security. Other potential additional indexes could be related to interest-rate movements and firm capitalization size. Sharpe (1984) finds quite a wide array of these additional factors, which he classifies as either a systematic influences or a sector influences. Sector influences includes eight factors: (1) basic industries, (2) capital goods, (3) construction, (4) consumer goods, (5) energy, (6) finance, (7) transportation, (8) utilities.

33. Sector Loadings

For correlation analysis, a firm or industry group is said to be dependent upon underlying economic factors or “sectors” such as: (a) the market as a whole, (b) interest rates, and (c) oil prices. As two industries “load” are influenced by common factors, they will have a higher correlation between them. This kind of loading is called sector loading.

34. Securitization

Pooling loans for various purposes into standardized securities backed by those loans, which can then be traded like any other security. In sum, the process of converting assets into marketable securities is called securitization.

35. Security

Collateral which a borrower pledges against a loan or secondary source of repayment in case of default.

36. Security Analysis

Security analysis is used to determine correct value of a security in the market-place. It is one of the three steps of forming portfolios of securities.

The selection of a portfolio of securities can be thought of as a multi-step process. The first step consists of studying the economic and social environment and the characteristics of individual companies in order to produce a set of forecasts of individual company variables. The second step consists of turning these forecasts of fundamental data about the corporation and its environment into a set of forecasts of

security prices and/or returns and risk measures. This step is often called the valuation process. The third and last step consists of forming portfolios of securities based on the forecast of security returns.

37. Security Characteristic Line

A plot of the excess return on a security over the risk-free rate as a function of the excess return on the market. [See **Market model**]

38. Security Dealers

Financial firms that provide a conduit for buyers and sellers of marketable securities by holding a portfolio of these securities and standing ready to buy and sell these securities at an announced price.

39. Security Interest

The legal claim on property that secures a debt or the performance of an obligation.

40. Security Market Line (SML)

A straight line that shows the equilibrium relationship between systematic risk and expected rates of return for individual securities. According to the SML, the excess return on a risky asset is equal to the excess return on the market portfolio multiplied by the beta coefficient. [See **Capital asset pricing model**]

41. Security Market Plane (SMP)

A plane that shows the equilibrium relationship between expected return and the beta coefficients of more than one factor.

42. Security Selection Decision

Choosing the particular securities to include in a portfolio. In general, the fundamental instead of technical analysis is used to select the particular stock to be in the portfolio.

43. Securities and Exchange Commission

Public firms must submit to regulation by the Securities and Exchange Commission (SEC) and the exchange on which their shares are traded. Regulation means more paperwork to file and more attorneys' fees to ensure that laws are not unintentionally broken. In the United States, public firms must file annual and quarterly reports with the SEC, and corporate insiders who buy and sell the firm's

stock must report their transactions to the SEC. The firm must register most public offerings of securities (including the initial public offering) with the SEC and receive SEC approval before selling the securities to the public.

Experts on public policy have known for some time that most employment growth in the United States comes from small businesses. To foster future growth, in 1992 the SEC adopted a series of rules to make it easier for small firms to raise public equity financing. The new regulations allow a firm to evaluate the potential market for its shares before committing to the time and expense of preparing a formal offering document.

Firms are allowed to raise up to \$1 million without registering the sale with the SEC. They can register securities worth up to \$5 million through the simpler and less costly Regulation A process rather than undergoing a full SEC review. The SEC estimates that the new regulations will reduce the cost of raising public equity by smaller firms by up to one-third. In addition, small public firms (by SEC definition, whose with sales less than \$25 million and market values of equity below \$25 million) will be able to file shorter, less comprehensive quarterly and annual SEC reports, thus reducing management's paperwork time and costs.

44. Seesaw Effect

When bonds are originally issued, most sell at prices close to par and offer coupon rates close to the market rates on bonds of similar maturity and risk. Over the life of a bond, its price will vary inversely to, or in the opposite direction of interest rate fluctuations in the economy. As interest rates rise in the economy, bond prices fall; as interest rates fall, bond prices rise. Since one rises as the other falls, we call this relationship between bond prices and interest rates the "seesaw effect."

45. Selection Phase

Because managers want to maximize the firm's value for the shareholders, they need some guidance as to the potential value of the investment projects. The selection phase involves measuring the value, or the return, of the project as well as estimating the risk and weighting the costs and benefits of each alternative to be able to select the project or projects that will increase the firm's value given a risk target.

46. Selective Credit Controls

Monetary policy tools that affect specific groups or sectors in the financial system.

47. Self- Financing Portfolio

A hedge portfolio that retains specified characteristics (e.g., it is zero-investment and risk-free) without the need for additional investments in the portfolio.

48. Self-Liquidating Loans

In view of high exposure to risk for a comparatively low return, commercial banks have understandably tried to find ways to protect themselves. Until very recently, this effort led them to lend only short-term funds and only in the form of self-liquidating loans – that is, they loaned money only for specific purposes and operations that would produce adequate cash flows to retire the debt quickly. The perfect example of such a self-liquidating situation is a working-capital loan made to a manufacturer or retailer that has a marked seasonal sales pattern.

For example, retail sales of a toy manufacturer's product peak just before Christmas each year. The manufacturer's own sales peak probably comes in August; however, when retailers and toy distributors are building up their inventories for the buying season, to meet this demand, the manufacturer must schedule a high level of production from May through July. In May of each year, therefore, the company takes out a loan from its bank to provide added working capital to finance the buildup in inventory. By September, heavy sales draw down the inventory to normal levels. Most of these sales, however, are made on terms of net 30 days, giving the company a large accounts receivable balance, but little cash. Finally, by early November, the customers pay their accounts, and collections of accounts receivable provide enough cash flow to retire the bank loan. Thus, the loan is self-liquidating in 6 months.

This is a classic bank lending situation. The bank knows before it makes the loan exactly how long the funds will be needed. The relatively short life of the loan increases the bank's liquidity. By making a fairly large number of predictable, short-term loans, a bank feels comfortable lending the highest proportion of its funds that regulations permit. In other words, it will want to lend up to its loan limit, or be fully loaned. If a bank finds little demand for self-liquidating, seasonal loans, it may be forced to lend in longer term, less predictable situations. Caution would probably lead this bank to keep a higher proportion of its funds in marketable securities to preserve its overall liquidity.

This traditional scenario has been transformed by important changes in bank practices during recent years. Commercial banks no longer stress the self-liquidating requirement as strongly as they once did. As the suppliers of short-term financing have become more competitive, banks have become more willing to provide longer term funds in the form of term loans. These new practices are creating an increasingly flexible source of short-term and intermediate-term funds for business organizations.

49. Sell Offs

[See **Voluntary restructuring**]

50. Sell-Side Analysts

The job of a sell-side research analyst is to follow a list of companies, all typically in the same industry, and provide regular research reports to the firm's clients. As

part of that process, the analyst will typically build models to project the firms' financial results, as well as speak with customers, suppliers, competitors, and other sources with knowledge of the industry.

51. Semidirect Finance

Any financial transaction (especially the borrowing and lending of money) that is assisted by a security broker or dealer.

52. Semistrong-Form Efficient Market

Different assumptions about information availability give rise to different types of market efficiency. [See also **Efficient market**]

In a semistrong-form efficient market, all *public* information, both past and current, is reflected in asset prices. The US stock market appears to be a fairly good example of a semistrong-form efficient market. For the most part, news about the economy or individual companies appears to produce quick stock price changes without subsequent trends or price reversals.

53. Seniority

The order of repayment; in the event of bankruptcy, senior debt must be repaid before subordinated debt receives any payment.

54. Sensitivity Analysis

Scenario analysis simultaneously modifies many variables that affect cash flows and net present value (NPV) to build different scenarios. [See also **Scenario analysis**] Sensitivity analysis changes one variable at a time from its base-case value; this isolates the effects on NPV of changes in individual variables. If large changes in NPV occur when the product price assumption or exchange rate assumption changes by, for example, 10%, then additional research may be warranted to better determine the likely market price or exchange rate. On the other hand, if NPV is relatively stable as the assumed salvage value changes, then great effort should not be expended in order to determine a more accurate estimate of salvage value.

One method for doing sensitivity analysis is to change each individual variable from its base-case value by some amount, say 5% or 10%, while holding all other variables constant at their base-case values. The resulting NPVs are computed and then recorded or graphed. A steep NPV graph indicates a variable that has a major impact on project success, especially as the NPV of the project is negative for some reasonable values of the variable. A more gently sloped NPV graph shows that a variable does not have a major influence on NPV, so additional research on likely values of this variable probably is not warranted. Spreadsheet packages allow sensitivity analysis to be done with ease.

Rather than arbitrarily changing each variable by some fixed percentage, the analyst might take a cue from scenario analysis and determine best-case and worst-case values for each variable. NPVs can be computed as each variable is adjusted to its best- and worst-case estimates while all other variables are held at their base-case amounts. This combination of sensitivity analysis can pinpoint which worst-case and best-case values affect NPV by the greatest amount.

55. Separation Property

The property that portfolio choice can be separated into two independent tasks: (1) determination of the optimal risky portfolio, which is a purely technical problem, and (2) the personal choice of the best mix of the risky portfolio and the risk-free asset.

56. Serial Bonds

A series of bonds offered by the same issuer with principal payments that are due at different maturities. Serial bonds are common for municipal bond issuers.

57. Serial Bond Issue

An issue of bonds with staggered maturity dates that spreads out the principal repayment burden over time.

58. Serial Correlation

[See **Autocorrelation**]

59. Serial Covariance

The covariance between a variable and the lagged value of the variable; the same as autocovariance.

60. Serialization

The splitting up of a single bond issue into several different maturities (used most often for state and local government bonds).

61. Series of Options

A series is a subset of a class and consists of all contracts of the same class (same asset) having the same expiration date and exercise price.

62. Service Charges

Fees imposed for bank services. They are small portion of operating income for all banks, but they are more important for small banks.

63. Service Corporation

A corporation formed by Saving and Loan association to conduct diversified line of business.

64. Service-Oriented Economy

An industrialized economy in which the production and delivery of services account for a growing share of national output and employment.

65. Set of Contracts Perspective

View of corporation as a set of contracting relationships among individuals who have conflicting objectives, such as shareholders or managers. The corporation is a legal contrivance that serves as the nexus for the contracting relationships.

66. Settlement

The time in a transaction at which all obligations of both the buyer and the seller are fulfilled.

67. Settlement Date

The actual exchange of one currency for another occurs on the settlement date. Settlement is handled by an association of 12 New York banks called CHIPS (Clearing House Interbank Payments System).

68. Settlement Price

A figure determined by the closing-price range that is used to calculate daily gains and losses in futures-market accounts (and thus margin calls) and invoice prices for deliveries.

69. Share Draft

Interest-bearing checking account offered by a credit union.

70. Shareholder

Holder of equity shares. The terms *shareholders* and *stockholders* usually refer to owners of common stock in a corporation.

71. Shareholder Wealth

Shareholder wealth is measurable and observable daily in the financial sections of newspapers (at least for publicly traded firms). Shareholder wealth is nothing more

than the market value of a firm's common stock. This market value of the shareholders' claim on a firm is equal to:

$$\begin{aligned}\text{Shareholder wealth} &= (\text{Common stock price}) \\ &\quad \times (\text{Number of common shares outstanding}).\end{aligned}$$

This relationship allows analysts to keep track of changes in shareholder wealth for competing firms in an industry on a regular basis to see which is most successful at returning value to its shareholders. As long as the number of common stock shares outstanding does not change appreciably, the market's perception of the firm and its management's actions will be reflected in the firm's stock price.

Many possible criteria can be used to evaluate firm performance. Total sales, the growth rate of the firm, market share, management's strategy, or earnings per share frequently are suggested as bases for evaluating and analyzing firms. We argue, however, that shareholder wealth is the best measure of firm performance for several reasons.

First, shareholder wealth is an objective, market-determined measure. It is not subject to manipulation (except in violation of securities laws) or limited to subjective evaluation. The value of a firm's stock is based on the joint decision of many investors, who generally are forward-looking and trade in markets that determine unbiased prices.

Second, any accounting performance measure (such as sales, earnings, or firm growth) is vulnerable to distortion by accounting principles, whose application may be somewhat subjective (such as when to recognize revenue or how quickly to depreciate assets). Rather than presenting an unbiased view of firm performance, accounting statements may be oriented toward the perspective that management wants to present. Additionally, accounting-based performance measures are always *historical*. They tell us where the firm has been. In contrast, shareholder wealth is a forward-looking measure) incorporating the market's objective assessment of the firm's prospects for the future.

Third, growth for the sake of growth or merely to increase market share may have dangerous long-run implications. It would be folly to invest scarce capital in plant, equipment, and workers with no plan for how these investments will pay for themselves and return capital (plus interest) to the financial markets. The firm may win the short-term market-share battle but lose the long-term war as poor returns on investments hurt its ability to raise capital, repay loans, and pay bills and workers. The firm may find it difficult to maintain its short-sighted market-share gains as leaner and more financially responsible rivals counterattack at the firm's weak points to reclaim their lost customers. Thus, sacrificing financial value for market share does not lead to successful, long-term business. Maximizing shareholder value does.

Fourth, a shareholder wealth orientation ensures adequate consideration of risk in a firm's decisions. A basic principle of finance is the risk/return trade-off. Higher risk investments must offer higher expected returns; otherwise, investors will purchase assets that offer the same return with less risk, or a higher return for the same risks. Projects that attempt to increase a firm's earnings or cash flow may favorably affect

the *size* and *timing* of cash flows, but they may ignore the *risk* component. Increases in a firm's exposure to risk, even in the face of rising earnings, may lead to a *lower* firm and shareholder value. Attempts to maximize returns without considering risk may harm the firm's long-run viability and value to shareholders, creditors, and employees.

Fifth, shareholder value is the best performance measure because it represents the firm's performance from the perspective of those who have their capital most at risk – namely, its owners. Bondholders or bankers are not owners, but creditors; they receive their interest and principal payments before shareholders receive any cash from the firm. Although a very important component of any successful firm, employees resemble creditors; in exchange for their labor, they receive income from the firm's owners. Overemphasizing customer quality, satisfaction, or service may lead to happy customers, but uncontrolled cost increases to meet these needs without sufficient revenues may ultimately lead to the firm's demise. Lenders, fearing nonpayment, may cut off lines of credit; employees, fearing cutbacks, may leave to take other jobs; and the financial market, upset with the firm's poor use of capital, may downgrade its value, harming its owners.

Focusing on shareholder value rectifies all of these potential problems. Smart managers make decisions to service customers in a cost-efficient manner. They treat and pay employees fairly; otherwise, unmotivated and unhappy employees become unproductive, cost – increasing employees who prevent the firm from satisfying its goals. By focusing on firm value, managers work to maintain stable relationships with financing sources so funds will be available to finance future growth or retrenchment needs.

Some may wonder whether focusing on shareholder wealth may be detrimental to the nonfinancial aspects of the firm. Quite the opposite, focusing on shareholder wealth is the best means of helping the long-term survival of the firm in a dynamic, global economy. Shareholder value simultaneously considers all of the influences on the firm; decisions that are made to maximize shareholder wealth reflect the best interests of all of the firm's constituents or stakeholders.

In particular, decisions to maximize shareholder wealth may benefit employees. Managers realize that a happy, stable work force both increases productivity and reduces costs. Human resource departments of firms often conduct studies to compare the benefits and costs of offering programs to meet workers' needs. For example, an on-site day care center for workers' preschool-aged children can benefit both the employees and shareholders by reducing employee absenteeism and job turnover. Similarly, innovations in flexible worker scheduling and career planning can add value to the firm by reducing employee turnover and the subsequent costs of hiring and training new workers. Wellness programs encourage healthy eating and exercise habits among employees and can reduce both absenteeism costs and health insurance premiums. Workers for some firms periodically leave the factory and visit customers to learn more about customers' needs. This results in better motivated workers who can appreciate the importance of their job to the customer.

Shareholder wealth as a measure of firm performance is objective and forward-looking, and it incorporates all influences on the firm and its stakeholders. No other

measure of firm performance is as inclusive and practical as a means for evaluating a firm's strategies.

72. Share-Equivalent

The position in shares that has equivalent dollar risk to a derivative. [See also **Delta**]

73. Shark Repellent

Shark repellents are anti-takeover amendments that firms add to their corporate charters to protect themselves from unfavorable takeovers. One such strategy is a supermajority rule that requires 95% of a firm's shareholders to approve a **tender offer**. [See also **Tender offer**] Another technique is a fair price amendment, which requires a suitor to acquire stock at essentially one price. To a certain extent, this protects the shareholders of the firm against two-tier acquisitions, in which the acquiring firm acquires one block of stock at a high price and then the remaining shares at a substantially lower price.

74. Sharpe Hedge Ratio

The Sharpe ratio is the asset management industry's go-to statistic for summarizing achieved (or back-tested) performance. It is the most-cited reason to hire or fire individual money managers. The relationship between risk and return is an essential concept in finance, and the ratio captures this in a single number by gauging how much return investors earned for each unit of risk.

75. Sharpe Ratio

For an asset, the ratio of the risk premium to the return standard deviation (i.e., $\frac{\bar{R}_i - R_f}{\sigma_i}$), where $\bar{R}_i$ is average rate of return for i th security or portfolio; R_f is the risk-free rate and σ_i is the standard deviation of rates of return for i th security or portfolio.

76. Sharpe's Measure

Reward-to-volatility ratio; ratio of portfolio excess return to standard deviation. [See also **Sharpe ratio**]

77. Shelf Life

Number of days it takes to get goods purchased and sold, or days in inventory.

78. Shelf Registration

The shelf registration process saves issuers both time and money. There is no cost or penalty for registering shelf securities and then not issuing them. Filing fees are

relatively low, and the firm can take some securities from the shelf and sell them immediately through one underwriter and then sell more later with another underwriter. This technique allows the issuer to determine which investment bank offers the best service.

Not every firm can use shelf registration. Firms must meet several size, credit quality, and ethics requirements:

1. The market value of the firm's common stock must be at least \$150 million.
2. The firm must have made no defaults on its debt in the previous 3 years.
3. The firm's debt must be investment grade (rated BBB or better).
4. The firm must not have been found guilty of violation of the Securities Exchange Act of 1934 in the previous 3 years.

Despite its attractiveness and lower cost, few firms have chosen to sell seasoned equity through shelf registration; one study found that only 5.6% of seasoned equity offerings were shelf registered. As this suggests, shelf registration does have several drawbacks for equity issues. The first is that the securities are sold with no prior due diligence and analysis by an investment bank. An investment bank may assist in selling shelf-registered securities; but its role is limited to marketing the shares, rather than serving as an independent, analytical third party. Consequently, the investment bank can provide little or no certification effect. In fact, the issuer can entirely bypass investment banks when selling shelf-registered stock; it can sell the shares directly to dealers and investors on a stock exchange.

Investors may see a second drawback since the firm can decide when to sell shelf-registered securities. A firm commitment underwriting may take several months. During that time, the firm is at risk of an adverse price move. A shelf offering imposes essentially no delay between submitting the short registration form and selling the shares. This gives the issuer the opportunity to wait for a run-up in the stock price before issuing shares, but smart investors may suspect a shelf sale of equity as a sign that the shares are overpriced.

For a third disadvantage, a shelf registration of common shares leads to uncertainty, which investors do not like. Investors view the shares sitting on the shelf as overhanging the market, ready to be sold at any moment. This potential supply of shares depresses prices and raises investors' concerns about opportunistic stock sales.

79. Shell Branches

Booking offices of multinational banks, usually set up offshore to attract deposits and avoid certain domestic banking regulations.

80. Short

A position is short with respect to a price if the position profits from a decrease in that price. A short-seller of a stock profits from a decrease in the stock price and,

hence, is short the stock. A seller of an option profits from a decrease in volatility and, hence, is short volatility.

81. Short Call

Selling a call (writing it) has risk-reward characteristics which are the inverse of the long call. However, one major distinction arises when writing calls (or puts) rather than buying them. That is, the writer can either own the underlying asset upon which he or she is selling the option (a **covered write**), or simply sell the option without owning the asset (**naked write**).

82. Short Call

Short call means an investor writes a call option to earn the option premium. The writer of short call options has an obligation to sell the stock to the owner of the call option at the exercise price.

83. Short Forward

The party to a forward contract who has an obligation to sell the underlying asset.

84. Short Hedge

Sale of a futures contract to protect against a price decline.

85. Short Position or Hedge

The sale of a futures contract in anticipation of a fall in the price of the underlying asset. Also obligates delivery of the commodity of financial instrument (and payment) if the position is left open to maturity.

86. Short Put

A put that has been sold (write a put) can be covered or uncovered.

87. Short Put

Short put means an investor writes a put option to earn the option premium. The writer of the put option has the obligation to buy the stock from the owner of the put option at the exercise price.

88. Short Rate

The interest rate applying for a very short period of time.

89. Short Rebate

The rate of return paid on collateral when shares are borrowed.

90. Short Run

That period of time in which certain equipment, resources, and commitments of them are fixed. [See **long run**]

91. Short Sales Allowed

In portfolio management, short sales are either allowed or not allowed. The procedure to calculate optimum weights is much more complicated if short sales allowed.

92. Short Straddle

A short straddle is an options strategy comprised of selling both a call option and a put option with the same strike price and expiration date. It is used when the trader believes the underlying asset will not move significantly higher or lower over the lives of the options contracts.

93. Short Vertical (Bear) Spread

A short vertical (bear) spread is an options strategy implemented by an investor who is mildly bearish and wants to maximize profit while minimizing losses. The goal is to net the investor a profit when the price of the underlying security declines. The strategy involves the simultaneous purchase and sale of either puts or calls for the same underlying contract with the same expiration date but at different strike prices.

94. Short-Against-the-Box

The short-sale of a stock that the short-seller owns. The result of a short-against-the-box is that the short-seller has both a long and short position and, hence, bears no risk from the stock yet receives the value of the share from the short sale.

95. Short-Sale

A transaction in which an investor borrows a security, sells it, and then returns it at a later date to the lender. If the security makes payments, the short-seller must make the same payments to the lender.

96. Short Squeeze

[See **Squeeze**]

97. Short-Run Operating Activities

Events and decisions concerning the short-term finance of a firm, such as how much inventory to order and whether to offer cash terms or credit terms to customers.

98. Short-Term Debt

An obligation having a maturity of 1 year or less from the date it was issued. Also called unfunded debt. It includes accounts payable, notes payable, etc.

99. Short-Term Risk-Free Rate

T-bill ratio is the short-term risk free rate and T-bond rate is the long-term risk-free rate.

100. Short-Term Securities

Securities that mature in 1 year or less. These securities include T-bill, commercial paper, and others.

101. Short-Term Tax Exempts

Short-term securities issued by states, municipalities, local housing agencies, and urban renewal agencies.

102. Shortage Costs

Costs that fall with increases in the level of investment in current assets.

103. Shout Option

A shout call option expiring at time T has the payoff $\max(0, S_{\hat{t}} - K, S_T - K)$, where $\hat{t}$ is the time and $S_{\hat{t}}$ is the price at which the option holder “shouted,” thereby guaranteeing an expiration payoff at least as great as $S_{\hat{t}} - K$. [See **Deferred-strike option**]

104. Side Effects

Effects of a proposed project on other parts of the firm. These effects can either be positive or negative.

105. Sight Draft

A sight draft is payable on presentation to the importer and the exporter usually receives the proceeds within 1–2 weeks. Normally, the exporter's bank (or its correspondent bank in the buyer's country) does not present the draft for payment by the importer until the merchandise has been delivered (usually by ship). When the importer (the buyer) has verified that all the paperwork is in order and that the goods have arrived, it pays the bank and receives title to the merchandise.

106. Signaling Approach

Approach to the determination of optimal capital structure asserting that insiders in a firm have information that the market does not; therefore the choice of capital structure by insiders can signal information to outsiders and change the value of the firm. This theory is also called the asymmetric information approach.

107. Sigma

[See **Vega**]

108. Simple Interest

Simple interest, unlike compound interest, pays a return only on the principal (the money originally invested) over successive periods. To calculate simple interest, multiply the principal by the interest rate, and multiply again by time. Simple interest does not include any compounding.

109. Simple Interest Method

A method of figuring the interest on a loan that charges interest only for the period of time that borrower actually has use of the borrowed funds.

110. Simple Linear Regression

Simple linear regression is a statistical technique that fits a straight line to a set of data points, thus providing an expression for a relationship between two variables. One of the more widely used regression techniques is the method of least squares. If x_i is the independent variable and y_i is the dependent variable, the linear equation

$$y_i = \alpha + \beta x_i + e_i$$

can be solved, where

- α = intercept
- β = slope of the least-squares line
- e_i = error term

The values α and β for n observations of x and y can be estimated as

$$\hat{\alpha} = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} = \frac{\text{Cov}(x, y)}{\text{Var}(x)}$$

in which $\text{Cov}(x, y)$ stands for the covariance – that is,

$$\sum_{i=1}^n \frac{(y_i - \bar{y})(x_i - \bar{x})}{n}$$

and $\text{Var}(x)$ stands for

$$\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$$

111. Simple Prospect

An investment opportunity where a certain initial wealth is placed at risk and only two outcomes are possible.

112. Simulation

[See **Monte Carlo simulation**]

113. Simulation Analysis

In reality, every variable relevant to the capital budgeting decision can be viewed as a random variable. **Scenario analysis** and **sensitivity analysis** limit the randomness aspects of each item by examining only a few values of each variable. [See also **scenario analysis** and **sensitivity analysis**] Simulation analysis attempts to realistically portray the relevant inputs to the capital budgeting project as random variables. Each variable, whether it be price, variable cost, project life, or some other item, is assumed to have a probability distribution with a known mean and variance.

In each simulation trial, computer analysis uses a random number generator to select values from each variable's probability distribution as the basis for net present value (NPV) calculation. This process is repeated many times; each time, numbers are randomly drawn from each probability distribution. After replicating the trials several thousand times, the statistical distribution of the computed NPVs is plotted, and the average NPV and its variance are computed. Unlike the NPV point estimates derived from scenario or sensitivity analysis, simulation analysis gives an estimated distribution of potential NPVs.

Of course, the simulation output is only as accurate as the inputs. It is likely that an inaccurate NPV distribution will result if inappropriate probability distributions, means, and variances are used as inputs.

114. Simulation Method for Capital Budgeting Decisions

Simulation is another approach to capital budgeting decision-making under uncertainty. In cases of uncertainty, every variable relevant to the capital budgeting decision can be viewed as random. With so many random variables, it may be difficult or impossible to obtain an optimal solution with an economic or financial model. Any model of a business-decision problem can be used as a simulation model if it replicates or simulates business problems and conditions. However, true simulation models are designed to generate alternatives rather than find an optimal solution. A decision is then made through examination of these alternative results. Another aspect of the simulation model is that it focuses on an operation of the firm in detail, either physical or financial, and studies the operations of such a system over time. Simulation is also a useful tool for looking at how the real system operates and showing the effects of the important variables. When uncertain, or random, variables play a key part in the operations of a system, the uncertainty will be included in the simulation and the model is referred to as a probabilistic simulation. Its counterpart, deterministic simulation, does not include uncertainty. Input variables for simulation generally include market size, selling price, market growth, market share, total required investment, useful life, residual value, operating cost, fixed costs, tax rate, and discount rate.

115. Single-Country Funds

Mutual funds that invest in securities of only one country. It is a closed-end fund.

116. Single-Factor Model

A model of security returns that acknowledges only one common factor. [See also **Factor model**]

117. Single Index Model

A model of stock returns that decomposes influences on returns into a systematic factor, as measured by the return on a broad market index, and firm-specific factors. This method can be used to simplify the portfolio selection process. [See **Market model**]

118. Single-Price Auction (Dutch Auction)

In an important experiment begun in 1992, the Fed instituted a single-price, or Dutch, auction system for selected 2-year and 5-year Treasury notes. In a traditional

Treasury auction, securities are allocated to the highest bidders, in descending order of the prices they bid, until all securities to be issued are awarded. Thus, winning bidders for the same security pay different amounts, and the highest bidder pay the same price. Many experts believe that the traditional bidding system encourages primary dealers, who must bid at every auction, to collude in their efforts to minimize the winner's curse. The traditional system may also encourage cornering: Winning bidders may attempt to compensate for the winner's curse by earning excess profits as they resell securities they have won at the auction.

119. Sinking Fund

It is a procedure that allows for the repayment of principal at maturity by calling for the bond issuer to repurchase some position of the outstanding bonds. An indenture may require the firm to retire specified portions of the bond issue over time through payments to a sinking fund. This provides for an orderly and steady retirement of debt over time. Sinking funds are more common in bonds issued by firm with lower credit ratings; a higher quality issuer may have only a small annual sinking fund obligation due to a perceived ability to repay investors' principal at maturity.

A sinking fund affects the maturity of a bond issue since it allows the firm to retire the issue in bits and pieces over time. After a deferral period following the primary market offering the sinking fund requirement usually can be satisfied in one of two ways. First, the issuer can select specific bonds for retirement by randomly drawing serial numbers. Investors whose numbers are drawn must return their bonds to the firm in exchange for repayment of principal. The issuer effectively calls in portions of the issue over time.

The second way to meet the sinking fund requirement is to purchase bonds from willing investors in the secondary market. Secondary market purchases become attractive if the bond's market price is less than par.

120. Size Effect

The portfolios of the firm with the smallest market value experienced return that were, both economically and statistically, significantly greater than the portfolios of the firm with large market value.

121. Skewness

A statistical measure which characterizes the asymmetry of a distribution around its mean. Positive skews indicate asymmetric tail extending toward positive values (right-hand side). Negative skewness implies asymmetry toward negative values (left-hand side). It is the third moment of a distribution. To calculate the sample skewness of random variable X :

$$S_x = \frac{N}{(N-1)(N-2)} \sum_{i=1}^N \left(\frac{X_i - \bar{x}}{\sigma_x} \right)^3,$$

where N is the number of observations; and $\bar{x}$ is the sample mean.

The distribution of losses across a credit portfolio will be positively skewed if there is positive correlation between obligors or the size/number of exposures is coarsely granular. This means that the confidence interval out on the downside tail will be further away from the mean than would be expected given the portfolio's standard deviation alone.

122. Skip-Day Settlement

A convention for calculating yield that assumes a T-bill sale is not settled until 2 days after quotation of the T-bill price.

123. Slack Variable

In an optimization problem, a slack variable is a variable that is added to an inequality constraint to transform it into an equality. Introducing a slack variable replaces an inequality constraint with an equality constraint and a non-negativity constraint on the slack variable.

124. Small Company Offering Registration

Several states offer programs to ease the process of public equity financing for firms within their borders. A firm in a state that has enacted a SCOR (Small Company Offering Registration) law can raise \$1 million by publicly selling shares worth at least \$5. This law creates a fairly standardized, fill-in-the blank registration document to reduce a firm's time and cost in preparing an offering. A firm can sell shares from a SCOR offering in other states with minimal notice to the SEC. According to estimates, a SCOR offering can reduce issuing costs by up to one-half for small firms.

125. Small-Firm Effect

Market anomaly whereby small companies exhibit a propensity to produce rates of return that are larger than those predicted on the basis of the capital asset pricing model.

126. Small Issues Exemption

Issues of less than \$50 million are governed by Regulation A. under the regulation A, only a brief offering statement instead of a lengthy regular statement is needed. Securities issues that involve less than \$1.5 million are not required to file

a registration statement with the Securities and Exchange Commission. For regulation A to be operative, no more than \$1.5 million may be sold by insiders.

127. Social Accounting

A system of recordkeeping which reports economic and financial activity for the whole economy and/or between the principal sectors of the economy, such as households and businesses.

128. Social Responsibility

A trend among financial institutions and their regulators to promote fair and equitable treatment of all customers and to make financial services available to all or most segments of the population.

129. Society for Worldwide Interbank Financial Telecommunications

The Society for Worldwide Interbank Financial Telecommunications (SWIFT) is not a settlement system, but a communications system that facilitates settlement of wire transfers through banks in different countries. Currently, over 1,600 banks maintain membership in the system, most of them located in the United States and Europe. The innovative feature of SWIFT is the standardization of messages so that computer software throughout the world can read SWIFT messages. The SWIFT network handles all types of customer and bank transfers.

130. Soft Dollars

The value of research services that brokerage houses supply to investment managers “free of charge” in exchange for the investment managers’ business.

131. Sole Proprietorship

A business owned by a single individual. The sole proprietorship pays no corporate income tax but has unlimited liability for business debts and obligation. [See **Proprietorship**]

132. Solicitation Method

A method for selling federal agency securities in which orders are taken from buyers and the securities are priced and then delivered to investors after the order book is closed.

133. Sources and Uses for Funds Statements

A financial report prepared for each sector of the economy in the Federal Reserve Boards’ Flow of Fund Accounts that shows changes in net worth and changes in holdings of financial assets and liabilities over a specific time period.

134. Sovereign Risk

[See **Country risk**]

135. Spark Spread

The difference between the price of electricity and that of the quantity of natural gas required to produce the electricity. Actually the operation cost to produce electricity includes not only gas cost. Therefore, the spark spread is the variable component of the marginal profit.

136. Special Drawing Rights (SDRs)

SDRs are a form of currency related by IMF in the 1970s to increase would liquidity. SDRs are a weighted average of the US dollar, the German mark, the Japanese yen, the French Franc, and British pound.

137. Specialist

Individual on the floor of an organized exchange who keeps an inventory of one or more stocks and trades with floor brokers out of that inventory.

138. Speculation

Undertaking a risky investment with the objective of earning a greater profit than an investment in a risk-free alternative (a risk premium). The distinction between hedging and speculation comes not from which side of futures contract one take but from the motivation for entering into the contract.

139. Speculative-Grade Bond

Bond rate Ba or lower by Moody's, or BB or lower by Standard and Poor's, or an unrated bond.

140. Speculator

A market participant who is willing (for a price) to take on the risk the hedger wishes to eliminate. This trader goes long or short on a contract without having, or intending to take, an opposite position in the cash market. Speculator can be either scalper or day trader. [See **Scalper**]

141. Spin-Off

In a spin-off, a parent firm distributes all shares in a wholly owned subsidiary to its shareholders, thus creating a new corporate entity (although the same owners have the same percentage of shares in the new firm). A spin-off may involve either

the stock of an existing subsidiary or newly created stock representing ownership in the disposed unit.

142. Spontaneous Financing

Short-term financial planning involves much more of the firm's operations than working capital management alone; it extends to management of all of the firm's current assets and current liabilities and their interrelationship. In practice, financial managers make little or no distinction between investment decisions involving current assets and financing decisions involving current liabilities. Current assets and current liabilities often are too closely related for such separate treatment. Both current asset and current liability accounts increase simultaneously, providing financing (at least in the short run) for the investment. For example, when the firm obtains inventory on credit, it generates an account payable. This is called spontaneous financing.

143. Spot Curve

The set of zero-coupon bond prices with different maturities, usually inferred from government bond prices.

144. Spot Exchange Rate

The spot exchange rate represents the exchange rate for an immediate exchange of two currencies. The actual exchange of one currency for another occurs on the settlement date, up to 2 days later.

145. Spot-Futures Parity Theorem

Describes the theoretically correct relationship between spot and futures prices. Violation of the parity relationship gives rise to arbitrage opportunities.

146. Spot Interest Rate

Interest rate fixed today on a loan that is made today. In addition, it can also refer to the interest rate on an investment that is made for a period time starting from today and last for n years.

147. Spot Market Transaction

Cash or spot market transaction represent the exchange of any asset between two parties who agree on the asset's characteristic and price, where the buyer tenders payment and take possession of the asset when the price is set. [See **Forward contract**]

148. Spot Price

The current price of the commodity if purchased in the cash or “spot” market.

149. Spot Rate

The current interest rate appropriate for discounting a cash flow of some given maturity.

150. Spot Trade

There are three types of trade that take place in the foreign exchange market: spot, forward and swap. Spot trade involves an agreement on the exchange rate today for settlement in 2 days.

151. Spot Volatilities

The volatilities used to price a cap when a different volatility is used for each caplet. [See **Flat volatility**]

152. Spread

Refers to the simultaneous purchase and sale of futures contracts for (1) the same commodity or instrument with different maturity months or (2) commodities in different but related markets.

153. Spread (Futures)

Taking a long position in a futures contract of one maturity and a short position in a contract of different maturity, both on the same commodity.

154. Spread (Options)

A combination of two or more call options or put options on the same stock with different exercise prices or times to expiration. A *money spread* refers to a spread with different exercise price; a *time spread* refers to differing expiration date.

155. Spread Underwriting

Difference between the underwriters’s buying price and the offering price. The spread is a fee for the service of the underwriting syndicate.

156. Spreadsheet

A computer program that organizes numerical data into rows and columns on a terminal screen, for calculating and making adjustments based on new data.

157. Squeeze

The possibility that enough long positions hold their contracts to maturity that supplies of the commodity are not adequate to cover all contracts. A *short squeeze* describes the reverse: short positions threaten to deliver an expensive-to-store commodity.

158. Stable Distribution

A probability distribution for which sums of random variables have the same distribution as the original random variable. Stable distribution can be classified as normal stable distribution and nonnormal stable distribution. The normal distribution is stable because sums of normally distributed random variables are stably distributed.

Nonnormal stable distribution can be further classified into (a) Symmetric stable distribution and (b) non-symmetric stable distribution. This kind of distribution can be used to describe the distribution of the real return of the stock prices.

159. Stable Dividend Policy

The Lintner (1995, pp. 49–95) study reinforces the notion that dividend policy conveys information to investors. Many financial managers strive to maintain steady or modestly growing dividends and avoid large fluctuations or changes in dividend policies. Reducing dividend fluctuations helps reduce investor uncertainty about future dividends. Lower risk leads to higher stock prices. Managers resist increasing dividends if they do not expect to maintain the increase in the future. This supports a predominant policy of maintaining historical dividends.

If firms hesitate to raise dividends too quickly, they positively abhor the prospect of reducing dividends, for several reasons. First, many individuals and institutions require large cash flows from their investments. For example, retired people in lower tax brackets generally covet high dividend payments. Tax-exempt institutions, such as endowment funds or pension funds, also need high current income and therefore desire high dividends. Miller and Modigliani argue that these individuals or institutions should ignore a stock's level of dividends because they always can liquidate some of their holdings in order to generate substantial transaction costs, especially brokerage fees. In addition to the time involved in deciding to sell securities, investors may exhaust all of their principal, leaving none for future income requirements.

Second, managers often resist reducing dividends also because a cut in dividends may be interpreted by the investment community as a signal of trouble with the firm

or a result of poor management. Even if the reduction is intended to allow the firm to pursue an attractive opportunity, it may adversely affect stock prices.

A third reason that firms resist reducing dividends involves the legal list. Many large, institutional investors are bound by the prudent man rule, or by legislation, to buy only securities that are included on the legal list. One criterion of the list is a long history of continued dividend payments without dividend reductions. Therefore, a firm that reduces or omits a dividend payment faces the risk of being ineligible for purchase by certain institutional investors.

A stable dividend policy can become a sort of self-fulfilling prophecy. An unexpected rise or reduction in dividends can have an announcement effect on the firm's share price. An increase in dividends may lead investors to perceive a promising future and share price may increase. A drop in dividends may lead investors to fear a less promising future, resulting in a drop in share price. These perceptions may be accurate if managers themselves feel it is important to avoid fluctuations, especially cuts. In such a company, investors would be correct in viewing dividend declarations as sources of information.

160. Stack and Roll

A hedging strategy in which an existing stack hedge with maturing futures contracts is replaced by a new stack hedge with longer dated futures contracts.

161. Stack Hedge

Hedging a stream of obligations by entering future contracts with a *single* maturity, with the number of contracts selected so that changes in the *present value* of the future obligations are offset by changes in the value of this "stack" of futures contracts.

162. Staggered-Maturity Plan

A common practice among bond-portfolio managers is to evenly space the maturity of their securities. Under the staggered-maturity plan bonds are held to maturity, at which time the principal is reinvested in another long-term maturity instrument. Little managerial expertise is required to maintain the portfolio and the maturing bonds and regular interest payments provide some liquidity.

163. Stakeholders

Both stockholders and bondholders are stakeholders of a firm.

164. Stand-Alone Principle

To properly estimate the cash flows of a proposed capital budgeting project, the project must be viewed in isolation from the rest of the firm. This stand-alone principle

ensures that analysts focus on the project's own cash flows, uncontaminated by cash flows from the firm's other activities.

165. Stand-Alone Percent Standard Deviation

Stand-alone standard deviation expressed as a percentage of the mean value for the given asset.

166. Standard and Poor's Bond Rating

AAA – bonds of highest quality

AA – high-quality debt obligations

A – Bonds that have a strong capacity to pay interest and principal but may be susceptible to adverse effects.

BBB – Bonds that have an adequate capacity to pay interest and principal, but are more vulnerable to adverse economic conditions or changing circumstances.

BB – bonds of lower medium grade with few desirable investment characteristics

B&CCC – primary speculative bonds with great uncertainties and major risk if exposed to adverse conditions

C – income bonds on which no interest is being paid

D – bond is in default

167. Standard and Poor's 500 Composite Index (S&P 500)

The S&P 500 index comprises industrial firms, utilities, transportation firms, and financial firms. Changes in the index are based on changes in the firm's total market value with respect to a base year. Currently, the base period (1941–1943 = 10) for the S&P 500 index is stated formally as follows:

$$S\&P\ 500\ index = \frac{\sum_{i=1}^{500} P_{ti} Q_{ti}}{\sum_{i=1}^{500} P_{0i} Q_{0i}} (10),$$

where:

P_{0i} = per-share stock price at base year 0

P_{ti} = per-share stock price at index data t

Q_{0i} = number of shares for firm i at base year 0

Q_{ti} = number of shares for firm i at index year t

The index is multiplied by an index set equal to 10. The specification of this index is identical to that of the value index indicated in the equation.

168. Standard Deviation

The standard deviation σ , is simply the square root of the **variance**. [See also **Variance**]

$$\sigma = \sqrt{\sigma^2}$$

The standard deviation formula gives units of measurement that match those of raw data.

Standard deviation can be given a statistical interpretation to help give the analyst an intuitive feel for the possible range of returns that can occur. If the underlying distribution of data is approximately normal, we expect 68% of the data terms to fall within one standard deviation of the mean, that is, $\bar{X} \pm \sigma$. About 95% of observed returns will fall within two standard deviations of the average $\bar{X} \pm 2\sigma$. Actual returns should fall within three standard deviations of the mean, $\bar{X} \pm 3\sigma$, about 99% of the time. Thus, if the mean and standard deviations are known, a rough range for future values can be estimated. [See also **Coefficient of variation**]

169. Standardized Normal Distribution

A normal distribution with an expected value of 0 and a standard deviation of 1. (Interval estimate for standardized normal distribution – see **Standard deviation**).

170. Standby Fee

Amount paid to an underwriter who agrees to purchase any stock that is not subscribed to the public investor in a rights offering.

171. Standby Underwriting

An agreement whereby an underwriter agrees to purchase any stock that is not purchased by the public investor.

172. Standstill Agreements

Contracts where the bidding firm in a takeover attempt agrees to limit its holdings of another firm. These agreements usually lead to cessation of takeover attempts and it has had a negative effect on stock prices.

173. State of the World

It is a credit rating migration outcome; a new credit rating arrived at the risk horizon. This can be either for a single obligor on a stand-alone basis or jointly between two obligors.

174. State-Chartered Banks

US banking corporations that received their charter of incorporation from a board or commission appointed by a governmental body in each of the 50 states.

175. Stated Annual Interest Rate

The interest rate expressed as a percentage per annum by which interest payment is determined. [See **Nominal interest rate**]

176. Stated Interest Rate

[See **Annual percentage rate**]

177. Statement of Cash Flows

The statement of cash flows (Financial Accounting Standards Board [FASB] Statement Number 95) can be derived using the balance sheets for two consecutive years and the most recent year's income statement. These inputs give the analyst insight into the firm's cash inflows and outflows; that is, they indicate how the firm raised and spent cash. It shows how income statement items and changes in balance sheet accounts affect the firm's cash position.

The statement of cash flows has three sections; cash flows from operating activities, cash flows from investing activities, and cash flows from financing activities. The sum of the cash flows from these three sections gives the net change in the cash position of the firm. In the language of the accountant, the items in this statement are "reconciled to cash."

The first step in constructing a statement of cash flows is to compute the change in each item between the beginning and ending balance sheets and to classify each as a source or a use of cash. Generally, a source of cash creates a cash inflow: a use of cash generates a cash outflow. A source of cash results when an asset account (except for the cash account) is *decreased* or when a liability or equity account is *increased*. Let us look at this intuitively. A reduction in accounts receivable implies that customers sent cash to the firm to pay their bills; a reduction in inventory implies that goods have been sold; a decline in fixed assets implies that assets have been sold for cash. Likewise, increases in accounts payable, notes payable, or debt figures imply that the firm has taken on additional financing sources; an increase in a common or preferred stock account implies that the firm has raised funds by a stock issue.

A use of cash leads to a cash outflow; a use of cash occurs when there is an increase in an asset account (except cash) or a reduction in a liability or equity account. Increases in inventory, or fixed assets, for example, imply that the firm used funds to purchase an asset. A reduction in a liability or equity account implies that the firm used cash to pay bills or repurchase securities.

In addition to balance sheet information, we also need the information from the income statement to construct a statement of cash flows. Generally, income is a source of funds and expenses represent a use of funds. However, non-cash expenses, such as depreciation, do not represent a cash outflow and are therefore not a use of funds.

Cash Flows from Operating Activities This section of the statement of cash flows lists the sources and uses of cash that arise from the normal operations of a firm. In general, the net cash flow from operations is computed as income statement net income plus adjustments for noncash revenues and expenses:

$$\begin{aligned}\text{Cash flow from operating activities} &= \text{Net income} + \text{Depreciation} \\ &\quad - \text{Change in modified net working capital}\end{aligned}$$

You may recall that net working capital is defined as the difference between current assets and current liabilities:

$$\text{Net working capital} = \text{Current assets} - \text{Current liabilities}$$

Thus, an *increase* in net working capital is a net investment in the firm's current assets; and an increase in an asset is considered a use of cash. A *decrease* in net working capital is a divestment of assets, that is, a source of cash.

A modified net working capital amount is used to compute cash flow from operating activities, as standard definitions of current assets include cash and marketable securities and standard definitions of current liabilities include notes payable. In the statement of cash flows, changes in notes payable are considered a financing flow and thus appear as a component of the cash flows from financing activities. The change in the cash account appears at the bottom of the statement, as the sum of cash flows from operating, investing, and financing activities

Cash Flows from Investing Activities This section of the statement of cash flows represents the investments a firm makes in both its own fixed assets and the equity of other firms, including subsidiaries or joint ventures. (These holdings are listed in the investment account of the balance sheet.) Increases and decreases in these accounts are considered investment activities. The cash flow from investment activities is the change in gross plant and equipment plus the change in the investment account. The changes are added if they represent a source of funds; otherwise, they are subtracted. The dollar changes in these accounts are computed from the beginning and ending balance sheets.

Cash Flows from Financing Activities This section of the statement of cash flows includes cash flows arising from purchases and sales of notes payable and long-term securities and dividend payments to equity holders (recall that interest payments to bond holders help determine the firm's net income, which is part of cash flows from

operating activities). Cash flows from financing activities are computed as financing sources minus financing uses. Sources include increases in notes payable and new issues of bonds, preferred stock, and common stock, since these actions result in cash inflows. Uses include principal payments or the repurchase of notes payable, bonds, or stock. Dividend payments to equity holders also are considered in financing use.

The sum of the cash flows from operating, investing, and financial activities is the net increase or decrease in the firm's cash. By detailing changes in important financial statement line items, the statement of cash flows reveals information that the balance sheet and income statement cannot provide.

178. Statewide Branching

Allowing banks to establish branches throughout an entire state.

179. Static CAPM

Static CAPM represents one period type of the capital asset pricing model. It assumes an investment opportunity set does not change over time.

180. Static Hedge

A hedge that does not have to be changed once it is initiated.

181. Static NPV

The net present value (NPV) of a project at a point in time, ignoring the possibility of postponing adoption of the project.

182. Static Option Replication

The use of options to hedge options, with the goal of creating a hedging portfolio that has a delta that naturally moves in tandem with the delta of the option being hedged (see Derman et al. [1995](#), pp. 78–95).

183. Static Theory of Capital Structure

Theory that the firm's capital structure is determined by a trade-off of the value of tax shields against the costs of bankruptcy.

184. Static Trade-Off Hypothesis

According to the static trade-off hypothesis, a firm balances the marginal benefits (tax shields) of additional debt financing with its marginal costs, namely the increase in the present value of future expected bankruptcy costs. Any

increases in debt beyond this optimal level actually reduce firm value, as investors' perceptions of the increased cost of bankruptcy outweigh the tax benefits of additional debt.

185. Statistical Distribution Method for Capital Budgeting Decisions

The statistical distribution method implies that we need to calculate the mean and standard deviation of NPV to estimate the interval for NPV in capital budgeting decisions. We can use either decision tree or simulation methods to estimate the mean and standard deviation of NPV.

186. Statutory Accounting

Statutory accounting is a combination of cash based and accrual accounting; expenses are recognized when paid but revenues are not recognized until earned. In general, it is a more conservative way of reporting final results than GAAP. Both thrifts and insurers use both generally accepted accounting principles and statutory accounting rule.

187. Step-Up Swap

A swap where the principal increases over time in a predetermined way.

188. Stochastic Differential Equation

An equation characterizing the change in a variable in which one or more of the differential terms are increments to stochastic process.

189. Stochastic Process

An equation describing the probabilistic behavior of a stochastic variable is called stochastic process. Stochastic processes can be classified as discrete time or continuous time. A discrete-time stochastic process is one where the value of the variable can change only at certain fixed points in time, whereas a continuous-time stochastic process is one where changes can take place at any time. Stochastic processes can also be classified as continuous variable or discrete variable. In a continuous-variable process, the underlying variable can take any value within a certain range, whereas in a discrete-variable process, only certain discrete values are possible.

190. Stochastic Variable

A variable whose future value is uncertain. A stochastic variable can be classified into either continuous or discrete variable. [See **Stochastic process**]

191. Stock Dividend

Managers can use stock dividends to change the firm's number of common shares outstanding. A stock dividend is a payout of dividends in the form of stock rather than cash. A stock dividend commonly is expressed as a percentage; for example, a 10% stock dividend means that a stockholder receives one new share for every ten shares currently owned.

192. Stock Exchanges

Secondary markets where already-issued securities are bought and sold by members. [See **Exchanges** and **Secondary market**]

193. Stock Index

An average of the prices of a group of stocks. A stock index can be a simple average of stock prices, in which case it is *equally weighted*, or it can be a weighted average, with the weights proportional to market capitalization, in which case it is *value-weighted*. [See **Stock market index**]

194. Stock Index Futures

Futures on a stock index. For example, S&P 500 futures and major index futures.

195. Stock Index Options

An option on a stock index. For example, S&P 500 options.

196. Stock Market Index

A stock market index is a statistical measure that shows how the prices of a group of stocks change over time. A stock market index encompasses either all or only a portion of stocks in its market. Stock market indexes employ different weighting schemes, so we can use this basis to categorize the indexes by type. The three most common types of stock market indexes are market-value-weighted indexes, price-weighted indexes, and equally weighted indexes. [See **Dow Jones Index**, **Standard and Poor's Index**, and **Wilshire 5000 Equity Index**]

197. Stock Options

One way to help solve the agency problem – help managers make decisions that are in shareholders' best interests is to relate the managers' personal wealth to shareholder value. Some firms tie managerial compensation to stock performance, often by awarding managers stock option as part of their compensation. The options allow managers to purchase, at a future time, a stated number of the firm's shares at a

specific price. If the firm's stock price rises, the value of the shares, and therefore the managers' wealth, also rises. Decisions that detract from the best interest of shareholders will affect management by making the stock options less valuable. More and more firms are basing the compensation of their top managers on the firm's stock price.

198. Stock Repurchase

A stock repurchase occurs when management spends corporate funds to buy back the stock of the company. A stock repurchase can benefit both management and shareholders. The repurchased shares become treasury stock and are then available for reissue to executives under stock option plans, to employees as part of profit sharing plans, and to other firms as part of mergers or acquisitions.

Management gains some defensive benefits by way of a stock repurchase. If managers of a cash-risk, low-debt firm fear a takeover, they may finance the stock purchase with the firm's excess cash or use debt, reducing the attractiveness of the firm as a takeover target. In addition, the repurchase program invites any dissatisfied stockholders to sell their shares back to the firm at a favorable price before a potential takeover company can make an offer for the stock. Management also benefits from the reduction in mailing a processing costs for annual reports, dividend payments, proxy statements, and other materials. Some repurchases are aimed directly at small shareholders for precisely this reason.

Shareholders may also benefit from a stock repurchase. Stockholders who want to sell their shares can do so at a favorable price. Stockholders who choose to hold onto their shares may benefit from the reduction in the number of shares outstanding. For example, suppose someone owns 1,000 shares of a company that has 25,000 shares outstanding. This stock represents a 4% ($1,000/25,000$) stake in the company. If the company repurchases 5,000 of its shares from other investors, then that stockholder's stake in the company increases to 5% ($1,000/20,000$).

Information related to cash dividends paid, repurchases of common stock, and employee compensation and stock option plans can be found in a firm's (consolidated) statement of common stock, retained earnings, and treasury stock.

199. Stock Selection

An active portfolio management technique that focuses on advantageous selection of particular stocks rather than on broad asset allocation choices.

200. Stock Split

Managers can use stock splits to change the firm's number of common shares outstanding. A stock split is essentially the same thing as a **stock dividend**, except that a split is expressed as a ratio instead of a percentage. [See also **Stock dividend**] Basically a stock dividend and a stock split increase the number of shares of stock outstanding without any cash flow to the firm or increase in firm value.

201. Stock-Index Arbitrage

The program trading strategy in which professional traders look for temporary underpricing or overpricing of stock-index financial futures contracts compared to the cash market (spot) prices of comparables stocks, simultaneously buying (selling) stocks and selling (buying) futures contracts to profit from any temporary relative mispricing of these two financial instruments.

202. Stockholder

Holder of equity shares in a firm. The terms *stockholder* and *shareholder* usually refer to owners of common stock.

203. Stockholders' Books

Set of books kept by firm management for its annual report that follows Financial Accounting Standard Board rules. The tax books follow the IRS rules.

204. Stockholders' Equity

The residual claims that stockholders have against a firm's assets, calculated by subtracting total liabilities from total assets; also net worth.

205. Stocks

Ownership shares in a corporation, giving the holder claim to any dividends distributed from current earnings.

206. Stop-Loss Order

A sell order to be executed if the price of the stock, which you already own, falls below a stipulated level. Order can also be differentiated on the basis of allowable time for completion.

207. Stop Payment

Request by a depositor to stop payment on a previously issued check that has not yet cleared.

208. Straddle

A straddle is a simultaneous position in both a call and a put on the same underlying asset. A long straddle involves purchasing both the call and the put. By combining these two seemingly opposing options an investor can get the best risk-return combination that each offers. A short straddle implies the position risk-return characteristics

of the long straddle. A short straddle is a simultaneous position in both a short call and a short put on the same underlying asset. Contrary to the long-straddle position, selling a straddle can be an effective strategy when an investor expects little or no movement in the price of the underlying asset. A similar interpretation of its use would be that the investor expects the future volatility of the underlying asset's price that is currently impounded in the option premiums to decline. Moreover, since the time decay is a positive effect for the value of this position, one appropriate time to set a short straddle might be in the last month to expiration for the combined call and put.

209. Straddle Rules

Tax regulations controlling the circumstances in which a loss on a claim can be realized when a tax payer continues to own related securities or derivatives.

210. Storage Costs

The costs of storing a commodity.

211. Straight Bond

A bond with no option features such as callability or convertibility.

212. Straight Voting

A shareholder may cast all of his or her votes for each candidate for the board of directors.

213. Straight-Line Depreciation

A method of depreciation whereby each year the firm depreciates a constant proportion of the initial investment less salvage value. [See also **Double-declining-balance method**]

For example, using the straight-line method, the firm can write off a uniform annual depreciation charge of \$1,080, a year, as shown below, when costs are \$6,000, and the salvage value is \$600:

$$\text{Annual depreciation} = \frac{\text{cost} - \text{salvage}}{\text{years}} = \frac{\$6,000 - \$600}{5} = \$1,080$$

214. Strangle

The purchase of a put and a call with the same time to expiration and different price. A strangle is a similar strategy to a straddle. The investor is betting that there will be a large price move but is uncertain whether it will be an increase or a decrease. [See **Straddle**]

215. Strap

A long position in two call options and one put option with the same strike price and expiration date. If a long position in one call and two put with same strike price and expiration date is called strip.

216. Strategic Planning

The process through which managers formulate the firm's mission and goals, and identify strengths, weaknesses, opportunities, and threats.

217. Stratified Sampling

A technique used in Monte Carlo valuation in which random numbers are drawn from each percentile (or other regular interval) of the distribution. [See **Quasi-random sequence**]

218. Street Name

Describes securities held by a broker on behalf of a client but registered in the name of the firm.

219. Strength Investing

Relative strength investing looks for stocks with better price movement so an investor can outperform the market. The idea behind relative strength investing is to find stocks that are outperforming the other stocks so you actually outperform the market. If the S&P 500 is up 10% in a year, for example, you want to use relative strength investing strategies to outperform the index. Otherwise, you can just buy the index, and that will get you diversification into various sectors.

220. Stress Testing

Testing of the impact of extreme market moves in the value of a portfolio. In addition to this, stress testing credit risk models imply to "back test" model to ascertain their predictive accuracy.

221. Striking Price

Price at which the put option or call option can be exercised. Also called the exercise price.

222. Strip

A variant of a straddle. A strip is two puts and one call on a stock, both with the same exercise price and expiration date. [See **Straddle**]

223. STRIPS

An acronym for *Separate Trading of Registered Interest and Principal of Securities*. STRIPS are the interest and principal payments from Treasury bonds and notes traded as individual securities. These securities were introduced by Merrill Lynch and Solomon Brothers in 1982.

224. Strip Hedge

Hedging a stream of obligations by offsetting each individual obligation with a futures contract matching the maturity and quantity of the obligation.

225. Stripped Bond

A bond in which individual coupon payments and principal payments are separated (stripped) from the bond and sold as distinct zero coupon securities.

226. Stripped of Coupons

Describes the practice of some investment banks that sell “synthetic” zero coupon bonds by marketing the rights to a single payment backed by a coupon-paying Treasury bond.

227. Stripped Securities

Securities that represent just the coupon interest or principal payments on a loan. The interest-only payment is referred to as an IO, while the principal-only payment is referred to as a PO.

228. Strong-Form Efficient Market

Different assumptions about information availability give rise to different types of market efficiency. [See also **Efficient market**]

A market in which prices reflect *all* public and privately available knowledge, including past and current information, is a strong-form efficient market. In such an efficient market, even corporate officers and other insiders cannot earn above-average, risk-adjusted profits from buying and selling stock; even their detailed, exclusive information already is reflected in current stock prices. Few markets can ever pass the test of strong-form efficiency. US laws prohibit *insider trading*, or trading based on important, nonpublic information. These laws reflect a public perception that it is unfair for someone with access to private information to use that position for their own profit. Remember that corporate officers should try to maximize shareholder wealth. Using inside information to benefit themselves at the expense of unknowing shareholders is a violation of the trust that should exist in the principal-agent relationship.

229. Structural Credit Risk Model

Structural and reduced form models represent the two primary classes of credit risk modeling approaches. The structural approach aims to provide an explicit relationship between default risk and capital structure, while the reduced form approach models credit defaults as exogenous events driven by a stochastic process (such as a Poisson jump process). Structural models, pioneered by Black, Scholes, and Merton, ingeniously employ modern option pricing theory in corporate debt valuation. Merton model was the first structural model and has served as the cornerstone for all other structural models.

230. Structured Note

A bond that makes payments that, at least in part, are contingent on some variable such as a stock price, interest rates, or exchange rates.

231. Subchapter S Corporation

A Subchapter S corporation is one of two special forms of corporate organizations in the United States that allow dividends to escape double taxation. A Subchapter S corporation (named for the section of the tax code that discusses this organization) must have fewer than 35 shareholders, none of which is another corporation. Income from a Subchapter S corporation flows untaxed to the shareholders; thus, it is taxed only once, as personal income of the shareholders.

232. Submartingale Model

A submartingale is a fair-game model where prices in the next period are expected to be greater than prices in the current period. A submartingale model is appropriate for an expanding economy. One with real economic growth, or an inflationary economy, one with nominal price increases.

233. Subordinated Debenture

A subordinated debenture is the riskiest type of bond. [See also **Debenture**] The claims of these bondholders are subordinate, or junior, to the claims of debenture holders. Most “junk bonds,” or high-yield bonds, are subordinated debentures.

234. Subordinated Debt

In the case of bankruptcy, the claims of holders of subordinated debt are subordinated to the claims of other debt holders. In banks, insured depositors are paid in full before holders of subordinated debt receive anything.

235. Subordination Clause

A provision in a bond indenture that restricts the issuer's future borrowing by subordinating the new lenders' claims on the firm to those of the existing bond holders. Claims of subordinated or junior debt holders are not paid until the prior debt is paid.

236. Subscription Price

Price that existing shareholders are allowed to pay for a share of stock in a rights offering. A rational shareholder will only subscribe to the rights offering if the subscription price is below the market price of the stock on the offer's expiration date.

237. Substitution Swap

Exchange of one bond for a bond with similar attributes but more attractively priced.

238. Sum-of-the-Year's-Digits Depreciation

Sum-of-the-year's-digits method is one of the accelerate depreciation methods. The annual depreciation of this method can be calculated as follows:

$$\text{depr}_t = \frac{N - (t - 1)}{\sum_{t=1}^N t} \times (\text{cost} - \text{salvage value}),$$

where depr_t = depreciation of the t th period and n = number of years. For example, if the equipment cost is \$6,000, and the salvage value is \$600, then the sum-of-years' digits for 5 years are determined as follows:

Year 1	$\frac{5-(1-1)}{1+2+3+4+5}$	×	(\$6,000-600)	=	\$1,800
Year 2	$\frac{5-(2-1)}{15}$	×	(\$5,400)	=	1,440
Year 3	$\frac{5-(3-1)}{15}$	×	(\$5,400)	=	1,080
Year 4	$\frac{5-(4-1)}{15}$	×	(\$5,400)	=	720
Year 5	$\frac{5-(5-1)}{15}$	×	(\$5,400)	=	360
					\$5,400

239. Sunk Cost

A sunk cost is a project-related expense that is not dependent upon whether or not the project is undertaken. For example, assume a firm commissioned and paid for a feasibility study for a project last year. The funds for the study have been spent

already; they represent a sunk cost. The study's cost is not an incremental cash flow as its cost is not affected by the firm's decision to either pursue or abandon the project. Therefore, the cost must be excluded from the project's cash flow estimates.

240. Super-Majority Amendment

A defensive tactic that requires 80% of shareholders to approve a merger.

241. Supervised Learning

Supervised learning is the most common subbranch of machine learning today. Typically, new machine learning practitioners will begin their journey with supervised learning algorithms. Supervised machine learning algorithms are designed to learn by example. The name "supervised" learning originates from the idea that training this type of algorithm is like having a teacher supervise the whole process. Supervised learning is frequently used to analyze the probability of credit card default.

242. Supply Shock

An event that influences production capacity and costs in the economy.

243. Supply-Side Economics

An approach to economic policy which argues that the nation's economic policy should be directed toward increasing productivity and the supply of goods and services in order to combat inflation.

244. Support Level

A price level below which it is supposedly difficult for a stock or stock index to fall.

245. Support Tranche

A class of mortgage-backed securities where the promised principal and interest payments are made after payments to holders of other classes of securities are made.

246. Surplus Funds

Cash flow available after payment of taxes in the project. [See **Free cash flow**]

247. Surplus-Budget Unit

An individual, business firm, or unit of government whose current income receipts exceed its current expenditures and therefore is a net lender of funds to the money and capital markets.

248. Sustainable Growth Rate

The sustainable growth rate measures how quickly the firm can grow when it uses both internal equity and debt financing to keep its capital structure constant over time. It is computed as follows:

$$\text{Sustainable growth rate} = \frac{(\text{RR})(\text{ROE})}{1 - (\text{RR})(\text{ROE})},$$

where RR is the firm's retention rate, which is multiplied by ROE, its return on equity, divided by one minus this product.

249. Swap

Exchange between two securities or currencies. One type of swap involves the sale (or purchase) of a foreign currency with a simultaneous agreement to repurchase (or sell) it. [See **Spot trade**]

250. Swap Agreements

Lines of credit established between central banks to be used when any particular central bank needs foreign currency to sell in the currency markets.

251. Swap Contract

In addition to using forward, futures, and option contracts to hedge transactions or transaction exposure, many corporations are engaging in what are called *swap transactions* to accomplish this. A swap contract is a private agreement between two companies to exchange a specific cash flow amount at a specific date in the future. If the specific cash flow amount is interest payments, then the contract is an **interest rate swap**; if the specific amount of cash flows is currency payments, then the contract is a **currency swap**. [See also **Interest rate swap** and **Currency swap**] The first swap contract was negotiated between IBM and the World Bank in the early eighties. Since that time, the swap market has grown to over \$10 trillion.

252. Swap Rate

The difference between the sale (purchase) price and the price to repurchase (resell) it in a swap. [See **Spot exchange rate** and **Forward exchange rate**]

253. Swap Spread

The difference between the fixed rate on an interest rate swap and yield on a Treasury bond with the same maturity.

254. Swap Tenor

The lifetime of a swap.

255. Swap Term

Another name for *swap tenor*.

256. Swapping

Swapping represents the process of swap. Swap can be classified as either interest rate swap or foreign exchange rate swap.

257. Swaption

Swaption represents option on swap. For example, an option to enter into an interest rate swap where a specified fixed bond rate is exchanged for floating-rate bond. Since the floating-rate bond is worth its face value at the start of a swap, swaption can be considered as options on the value of fixed-rate bond with strike price equal to the face value.

258. Swing Option

Swing option is also called take-and-pay option. It is an option created by trading the underlying asset. For example, energy option is which the rate of consumption must be between a minimum and maximum level. There is usually a limit on the number of time the option holder can change the rate at which the energy is consumed.

259. SWOT Analysis

SWOT analysis examines a firm's strengths, weaknesses, opportunities, and threats. It can help managers identify capital budgeting projects that will allow the firm to exploit its competitive advantages or prevent others from exploiting its weaknesses. Strengths and weaknesses arise from the firm's internal abilities, or lack thereof. Opportunities and threats represent external conditions that affect the firm, such as competitive forces, new technologies, government regulations, and domestic and international economic trends.

Strengths give the firm a comparative advantage in the marketplace. Perceived strengths can include good customer service, high-quality products, strong brand image, customer loyalty, innovative R&D efforts, market leadership, and strong financial resources. Managers must continue to develop, maintain, and defend these strengths through prudent capital investment policies or else they will diminish and shareholder wealth will decline as new and existing competitors take advantage of the weakening firm.

A firm's weaknesses give its competitors the opportunity to gain advantages over the firm. Once weakness is identified, the firm should select capital investments to

mitigate or correct them. For example, a domestic producer in a global market can try to achieve global economies of scale (i.e., achieve “global scale”) by making investments that will allow it to export or produce its product overseas. Such a move also may make it easier for the firm to raise money in the future, as it may be able to raise funds in several different financial markets instead of just in its home country.

260. Symbiotic

A conglomerate financial firm that frequently merges insurance sales, security brokerage, real estate brokerage, financial counseling, and credit services within the same organization.

261. Syndicates

For most firm commitment underwritings, the managing investment bank arranges investment banking syndicates to help distribute shares of the newly public firm. Syndicates serve several purposes. First, a syndicate broadens the market base to include clients from other investment banking firms, thus allowing a broader distribution of the new issue. Second, the syndicate allows the managing investment bank to diversify or spread the risk of underwriting the new issue. Rather than purchasing the entire issue, the managing investment bank actually commits capital to purchase and resell only a portion of the issue; the remainder of the funds comes from members of the syndicate.

262. Syndicated Loan

A loan provided by a group of financial institutions (FIs) as opposed to a single lender. A syndicated loan is provided structured by the lead FI (or agent) and the borrower once the terms (rates fees and covenants) are set, pieces of the loan are sold to other FIs.

263. Synergies

Synergy is the interaction or cooperation of two or more organizations, substances, or other agents to produce a combined effect greater than the sum of their separate effects. Financial synergy is when the combination of two firms together results in greater value than if they were to operate separately. Financial synergies are most often evaluated in the context of mergers and acquisitions. These types of synergies relate to improvement in the financial metric of a combined business such as revenue, debt capacity, cost of capital, and profitability.

264. Synthetic Option

Rubinstein and Leland (1981) suggest a strategy that replicates the returns on a call option by continuously adjusting a portfolio consisting a stock and a risk-free

asset (T-bill, cash). This is called a synthetic call-option strategy; it involves increasing the investment in stock by borrowing when the value of stocks is increasing, and selling stock and paying off borrowing or investing in the risk-free asset when market values are falling.

The key variable in this strategy is the delta value, which measures the change in the price of a call option with respect to the change in the value of the portfolio of risky stocks. For deep-in-the-money options, the delta value will be close to one because a \$1 change in the stock value will result in approximately a \$1 change in the option value. Thus to replicate the option with cash and stock, almost one share must be purchased and the amount borrowed will be approximately equal to the exercise price. For deep out-of-the-money options, the value of the delta will be close to zero, and the replicating portfolio will contain very few shares and little or no borrowing. Hence in its simplest form the delta value largely depends on the relationship between the exercise price and the stock price. As the market moves to new levels, the value of the delta will change; hence the synthetic option portfolio must be rebalanced periodically to maintain the proper mix between equity and borrowing or cash.

In a similar manner, a portfolio manager can create replicated put options through a combination of selling short the asset and lending. The amount of stock sold short is equal to the delta value minus one. As the market decreases in value, more of the equity is sold (the short position increases), with the proceeds invested at the risk-free rate. If the market increases in value, money is borrowed to buy the stock and reduce the short position.

265. Systematic Influences

Systematic influences includes

1. Beta (the slope of the regression of excess return for the security against excess return on the S&P index)
2. Dividend yield
3. Size
4. Bond beta
5. Alpha

[See **Sector influences**]

266. Systematic Risk

Diversification cannot eliminate risk that is inherent in the macro-economy; this risk is called systematic or **market risk**. General financial market trends affect most companies in similar ways. Macroeconomic events, such as changes in GDP, rising optimism or pessimism among investors, tax increases or cuts, or a stronger or weaker dollar have broad effects on product and financial markets. Even a well-diversified portfolio cannot escape these effects.

The only risk that should matter to financial markets is an asset's systematic, or market risk – that is, the sensitivity of the asset's returns to macroeconomic events. The **unsystematic**, microeconomic component of an asset's total risk disappears in a well-diversified portfolio. [See also **Unsystematic risk**] When financial markets evaluate the trade-off between risk and expected return, they really focus on the trade-off between *systematic risk* and expected return.

Systematic risk (or market risk) is the risk that is inherent in the system. As such, it cannot be diversified away. The only way to escape systematic risk is to invest in a risk-free security. A risk-free asset, by definition, will have no systematic risk. In sum, only the systematic portion of risk matters in large, well-diversified portfolios. Thus, the expected returns must be related only to systematic risk.

1.20 T

1. T-Bill

T-bill has a short time horizon and the backing of the US government, which gives it an aura of safety. T-bills, T-notes, and T-bonds are liquid assets. Therefore, they are marketable securities which are parts of current assets. [See **Treasury bills**]

2. T-Period Holding-Period Return

The percentage return over the T-year period an investment lasts.

3. TAC

Targeted amortization class (TAC) mortgage-backed securities in which payments are guaranteed for one specific prepayment rate.

4. TAIEX Options

TAIEX represents Taiwan Exchange Generalization Weight Stock Index. TAIEX options are the options on the TAIEX index futures. The underlying assets of TAIEX options are futures with no restriction on short sales.

5. Tail VaR

The expected loss conditional upon the VaR loss being exceeded. [See **Value at risk**]

6. Tailing

A reduction in the quantity of an asset held in order to offset future income received by the asset.

7. Take-and-Pay Option

[See **Swing Option**]

8. Takedown Risk

In making the loan commitment, the financial institution must always stand ready to provide the maximum of commitment line. The borrower has the flexible option to borrow anything between \$0 and the commitment amount (\$5 million for example) on any business day in the commit period.

9. Takeover

General term referring to transfer of control of a firm from one group of shareholders to another. Takeover can occur by acquisition, proxy contests, and giving-private transaction. Thus, takeover encompasses a broader set of activities than acquisitions.

10. Taking Delivery

Refers to the buyer's actually assuming possession from the seller of the asset agreed upon in a forward contract.

11. Tangible Equity

Total assets minus intangible assets minus total liabilities. In a bank the largest intangible asset is goodwill which represents dollar values that may not be realized should the combined institution from merger be forced to liquidate.

12. Target Cash Balance

Optimal amount of cash for a firm to hold, considering the trade-off between the opportunity costs of holding too much cash and the trading costs of holding too little. [See **Optimal cash balance**]

13. Target Firm

A firm that is the object of a takeover by another firm.

14. Target Payout Ratio

A firm's long-run dividend-to-earnings ratio. The firm's policy is to attempt to pay out a certain percentage of earnings, but it pays a stated dollar dividend and adjusts it to the target as increases in earnings occur. [See **Lintner's model**]

15. Targeted Repurchase

The firm buys back its own stock from a potential bidder, usually at a substantial premium, to forestall a takeover attempt. This kind of offer will not extend to other shareholders.

16. Tax Anticipation Notes

Short-term municipal debt to raise funds to pay for expenses before actual collection of taxes.

17. Tax Books

Set of books kept by firm management for the IRS that follows IRS rules. The stockholders' books follow Financial Accounting Standards Board (FASB) rules.

18. Tax Credit

Direct reduction in tax liability arising from qualifying expenditures.

19. Tax Deferral Option

The feature of the US Internal Revenue Code that the capital gains tax on an asset is payable only when the gain is realized by selling the asset.

20. Tax Reform Act of 1986

A legislated revision of the US federal income tax code lowering individuals' tax rates and raising taxes on corporations as well as simplifying the federal tax code to make it easier to understand and to reduce the importance of tax laws in making economic and financial decisions.

21. Tax Swap

Swapping two similar bonds to receive a tax benefit.

22. Taxable Acquisition

An acquisition in which shareholders of the acquired firm will realize capital gains or losses that will be taxed.

23. Taxable Income

Gross income less a set of deductions. These deductions are expenses items as presented in income statement. It is called earnings before tax.

24. Tax-Deferred Retirement Plans

Employer-sponsored and other plans that allow contributions and earnings to be made and accumulate tax free until they are paid out as benefits.

25. Tax-Equivalent Yield

Tax-exempt interest yield converted to a pretax taxable equivalent by dividing the nominal rate by 1 minus the investor's marginal income tax rate.

26. Tax-Exempt Securities

Debt securities issued by state, city, county, and other local units of government or by other qualified borrowers whose interest income is exempt from federal taxation and from most state taxes as well.

27. Tax-Exemption Privilege

A feature bestowed by law on some financial assets (such as state and local government bonds) that make the income they generate free of taxation at federal or state and local government levels, or both.

28. Tax-Exempts

Tax-exempts are debt obligations of municipalities, states, and federal agencies like the Public Housing Authority. The interest they pay is tax-free. When purchased close to their maturities, these issues have risk characteristics similar to those of other government securities, which depend, of course, on the creditworthiness of the issuers.

29. Tax-Free Acquisition

An acquisition in which the selling shareholders are considered to have exchanged their old shares for new ones of equal value, and in which they have experienced no capital gains or losses.

30. Tax-Timing Option

Describes the investor's ability to shift the realization of investment gains or losses and their tax implications from one period to another.

31. Technical Analysis

Research to identify mispriced securities that focuses on recurrent and predictable stock price patterns. This analysis does not consider the fundamental variables which are considered by fundamental analysis.

32. Technical Analysis

Technical analysis is a trading discipline employed to evaluate investments and identify trading opportunities by analyzing statistical trends gathered from trading activity, such as price movement and volume. Technical analysis can be classified as traditional method and machine learning method.

33. Technical Insolvency

Technical insolvency is the inability of the firm to meet cash payments on contractual obligations. The lack of cash to meet accounts payable, wages, taxes, interest, and debt retirement will constitute technical insolvency, even if the enterprise has adequate assets and generates both economic and accounting profits.

When assets are plentiful in relation to liabilities, a financial manager usually can plan ahead and arrange for sufficient cash through various sources to prevent any embarrassment. Most liquidity problems can be overcome by borrowing or through the planned liquidation of certain assets. A sound, profitable business should have no difficulty in this regard, and reasonable intelligent planning should ward off the danger of technical insolvency. However, if the firm is technically insolvent because of successive losses, poor management, or insufficient investment in working capital, then lenders will be less willing to place funds at its disposal.

The financial manager also should be aware of the potential for variability in the availability of funds. Even a willing lender often is hesitant during periods of tight money, great financial uncertainty, or panic.

34. Technicians

They believe past price change can be used to predict future price movements. Technicians interested in aggregate market forecasting would obviously want to examine past movement of different market indicator series. [See also **Chartists**]

35. Technology and Operation Risks

Technology and operational risks are closely related and in recent years have caused great concern to Financial Institution managers and regulators alike. The Bank for International Settlements (BIS), the principal organization of Central Banks in the major economies of the world, defines operational risk (inclusive of technological risk) as “the risk of direct or indirect loss resulting from inadequate or failed internal processes, people, and systems or from external events.” A number of FIs add reputational risk and strategic risk (e.g., due to a failed merger) as part of a broader definition of operational risk.

36. TED (Treasury Eurodollar) Spread

The difference between the 3-month Eurodollar rate and 3-month Treasury rate.

37. Temporary Working Capital

Some working capital needs persist over time, regardless of seasonal or cyclical variations in sales. In contrast to **permanent working capital**, temporary working capital consists of the additional funds required to meet the seasonal or cyclical variations in sales, over and above permanent working capital. [See **Permanent working capital**]

38. Tender Offer

In a tender offer, the acquiring firm makes its offer directly to the shareholders of the firm it wishes to acquire. This usually is accomplished through the financial press. The acquiring firm offers to pay a fixed amount per share to each shareholder who tenders shares; this price usually is set far enough above the current market price to entice the shareholders of the target firm.

Tender offers can be made when negotiation breaks down or as a surprise move by one firm to catch the management of the other firm off guard. A tender offer may bid either cash or stock, or some combination, for a block of shares of the target firm. In many large corporations, effective management control can be gained with ownership of less than 50% of the shares. Hence, an acquiring firm can make a tender offer, gain control, and then proceed to negotiate for the remainder of the shares.

State and federal laws impose several legal requirements on tender offers. A bid for shares must remain open for at least 20 days. Moreover, shares that are tendered during this period may be withdrawn during the period. If the bidder raises the original offer price, shares that were tendered under the original offer also are entitled to the higher price. After one firm makes a tender offer, other firms may join the battle for a target firm.

Tender offers have generated a new and colorful vernacular. Some of the frequently encountered terms are **white knight**, **shark repellent**, **pac-man strategy**, and **golden parachute**. [See also **White knight**, **Shark repellent**, **Pac-man strategy**, and **Golden parachute**]

39. Tenor

Time to maturity or expiration of a contract, frequently used when referring to swaps.

40. Term Bonds

Most bonds are term bonds, which mature at some definite point in time. Thus,

$$PV = \sum_{t=1}^n \frac{I_t}{(1 + k_b)^t} + \frac{P_n}{(1 + k_b)^n}$$

where:

I_t = the annual coupon interest payment

P_n = the principal amount (face value) of the bond

n = the number of periods to maturity

41. Term Insurance Policy

It provides a death benefit only, no build-up of cash value for a specified period. The most popular type of term life policy involves a premium that increases with age for a constant amount of death benefits.

42. Term Loan

Loan with a maturity beyond 1 year, typically repaid from the borrower's future cash flow. Interest and principal are repaid at maturity, and the lender must take a more active role in checking the borrower's compliance.

43. Term Premiums

Excess of the yields to maturity on long-term bonds over those of short-term bonds. [See **Liquidity premium**]

44. Term Repo

A repurchase agreement lasting for a specified period of time longer than 1 day.

45. Term RPs

Repurchase agreements (RPs or REPOs) with maturity beyond 1 day. RPs involve a loan between two parties with one either a securities dealer or commercial bank. [See **Term repo**]

46. Term Structure of Interest Rates

The term structure of interest rates arises from the risk-return relationship among debt securities. The term structure of interest rates is typically described by the yield curve. Typical yield curve diagrams use data for Treasury securities to eliminate risk of default from the analysis. However, similar curves can be constructed using corporate bonds of different maturities with the same credit rating. Over time, the term structure shifts upward or downward and becomes steeper or flatter, depending upon market influences on short-term and long-term interest rates. The term structure generally slopes upward, showing that long-term debt must offer investors a higher return (and borrowers a higher cost) than short-term debt.

The current shape and expected future changes in the shape of the term structure affect the firm's debt financing decision. Low long-term interest rates may convince treasurers to issue long-term debt to lock in low financing costs while they can. As the term structure becomes steeper, however, the temptation rises to issue short-term debt and simply sell new short-term debt issues to replace maturing ones (i.e., to roll over maturing short-term debt).

Favoring short-term debt over long-term financing can generate enormous cost savings, boosting the firm's profitability and marketing efforts.

47. Terminal Value

The value at maturity.

48. Terms of Sale

Conditions on which firm proposes to sell its goods and services for cash or credit.

49. Theta

The change in the value in the value of a derivative solely due to the passage of time. Based upon the call option formula defined in option pricing model. [See **Option pricing model**] The mathematical result can be defined as:

$$\frac{\partial C}{\partial T} = \frac{S\sigma}{2\sqrt{T}}N'(d_1) + rXe^{-rT}N(d_2) > 0$$

50. Third Market

Trading of exchange-listed securities on the OTC market.

51. Third-Country Bills

Bankers' acceptances issued by banks in one country that finance the transport or storage of goods traded between two other countries.

52. Thrifts

They include savings and loan associations, savings banks, and mutual savings banks. These institutions traditionally rely upon savings deposits as sources of funds. Hence, they also call savings institutions. However, they are now able to offer checkable deposit.

53. Tick

Refers to a change in price, either up or down. The amount varies with each contract.

54. Time Decay

Another term for theta. [See **Theta**]

55. Time Drafts

Time drafts are payable after a period of time. These drafts specify that payment is required in 30, 60, or 90 days, or more. A time draft allows the buyer to take title to the merchandise when it promises to pay, rather than when it actually pays the draft.

56. Time Series Analysis of Financial Ratios

Financial ratios can be used in *time series analysis* to evaluate firm performance over time. The best information source for time series analysis of firm ratios is the firm's financial statements and their footnotes. These materials appear in annual reports as well as 10-Q and 10-K filing with the Securities and Exchange Commission.

57. Time Spread

[See **Spread (options)**]

58. Time Value (of an Option)

The part of the value of an option that is due to its positive time to expiration. Not to be confused with present value or the time value of money. [See **Theta**]

59. Time Value of Money

The time value of money is one of the most important concepts in finance. The time value of money means that a dollar today is worth more than a dollar at any time in the future. The time value of money is a basic building block for much financial analysis. Proper decisions depend upon comparing present cash flows with cash flows in the distant future.

To evaluate the time value of money, four key concepts must be understood: (1) the future value of a single sum; (2) the present value of a single sum; (3) the future value of an annuity; and (4) the present value of an annuity

60. Time-Weighted Return

An average of the period-by-period holding-period returns of an investment.

61. Times Interest Earned

[See **Interest coverage ratio** and **Capital structure ratios**]

62. Timing Adjustment

Adjustment made to the forward value of a variable to allow for the timing of a payoff from a derivative.

63. TINSTAFL Principle

Economics teaches the TINSTAFL principle: “There is no such thing as a free lunch.” Capital budgeting analysis frequently applies this principle to existing assets.

64. Tobin’s Q

Market value of assets divided by replacement value of assets. A Tobin’s Q ratio greater than 1 indicates the firm has done well with its investment decisions. It can be approximated by market value/book value ratio.

65. Tombstone

Tombstone is an advertisement that publicizes a security offering. They are placed in newspapers and magazines by the managing investment bank to advertise its role in forming a syndicate and helping distribute the new issue. The ad lists the name of the issuer, the type of security issued, the quantity sold, the offering price, and the members of the investment banking syndicate. The managing investment bank is listed first, for a particularly large or lucrative offering, two or more managing investment banks may share the top position. Members of the syndicate are listed below, in different tiers, with the firms in each tier typically listed in alphabetical order. The most prestigious investment banks are listed in higher tiers; less prestigious banks appear in the lower alphabetized lists. Investment banks sometimes quarrel, and some even have pulled out of deals, over objections to the placement of their name in the tombstone ad. Firms in a tier that are listed out of alphabetical order are less prestigious members of that tier.

66. Total Asset Turnover Ratio

More efficient use of assets to generate sales boosts a firm’s growth rate, if all else remains constant. A higher turnover allows the firm to increase sales without a large increase in assets. [See also **Asset management ratios**]

67. Total Cash Flow of the Firm

Total cash inflow minus total cash outflow.

68. Total-Debt-to-Total-Assets Ratio

[See **Leverage ratios**]

69. Total Return Swap

A swap where the return on an asset such as a bond is exchanged for LIBOR plus a spread. The return on the asset includes income such as coupons and the change in value of the asset.

70. Total Risk

Total risk is defined as the sum of systematic and unsystematic risk. Total risk is also equal to the sum of all of the risk components. However, the importance and the contribution to total risk depend on the type of security under consideration. The total risk of bonds contains a much larger fraction of interest-rate risk than the total risk of a stock.

71. Tracking Problem

A perfect hedge is usually not possible because the correlation between the market index and the portfolio may not be perfect. This is called the tracking problem. The greater the correlation between the portfolio and the index, the more effective the hedge, the lower the correlation, the less effective the hedge as a portfolio insurance strategy.

72. Trade Acceptance

Written demand that has been accepted by a firm to pay a given sum of money at a future date.

73. Trade Barrier

Trade barriers reduce import quantities. They prevent domestic consumers from buying all of the foreign goods that they otherwise might buy. This reduced demand for foreign goods reduces demand for foreign currencies; thus, trade barriers can strengthen the currency of the country that erects them. Should two countries place trade barriers against each other, however, their effects may offset one another, with a net impact on exchange rates of zero.

74. Trade Credit (Receivables)

The balance sheet of any company lists accounts receivable on the asset side and accounts payable on the liability side. These categories represent credit extended to other companies (accounts receivable) and credit extended by other companies (accounts payable). These line items measure trade credit, a form of short-term financing provided by a selling company to a buying company. Essentially, the seller provides a loan to the buyer by allowing the buyer to postpone payment while taking immediate possession of goods or services.

The selling company can increase overall sales through trade credit, but now without cost. The decision to extend trade credit depends upon the incremental gain per unit of additional risk. Because many firms tie up a lot of capital or assume large obligations in trade credit transactions, decisions involving the management of credit can have a significant impact on cash flow, cost of capital sales growth, and debt capacity.

Trade credit is one of those decisions that affect all aspects of the firm – marketing, production, finance, and so on. Each of these functional areas will have a distinct view of the role of trade credit.

75. Trading Account

Securities debt and equity securities that are bought and hold primarily for the purpose of selling or trading in the near term. Institutions serving as major dealers in money market assets must keep an inventory of trading account securities from which to make trade with customers.

76. Trading Costs

Costs of selling marketable securities and borrowing.

77. Trading Range

Price range between highest and lowest prices at which a security is traded. In statistics, it is called range. This measure can be used to measure the variability of a random variable.

78. Trading Strategies

In finance, a trading strategy is a fixed plan that is designed to achieve a profitable return by going long or short in markets. The main reasons that a properly researched trading strategy helps are its verifiability, quantifiability, consistency, and objectivity. For every trading strategy one needs to define assets to trade, entry/exit points and money management rules. Bad money management can make a potentially profitable strategy unprofitable.

79. Trading Volume

Many technical analysts believe that it is possible to detect whether the market in general and/or certain security issues are bullish or bearish by studying the volume of trading. Volume is supposed to be a measure of the intensity of investors' emotions. If high volume occurs on days when prices move up, the overall nature of the market is considered to be bullish. If the high volume occurs on days when prices are falling, this- is a bearish sign.

80. Tranche

The principal amount related to a specific class of stated maturities on a collateralized mortgage obligation. [See **Collateralized mortgage obligation**]

81. Transaction Cash

A company's cash needs fall into three categories: (1) cash for day-to-day transactions, (2) reserve cash to meet contingencies, and (3) cash for compensating balance requirements. The required level of day-to-day transaction cash depends upon the number, frequency, and amount of anticipated transactions. The only requirement for this element of the cash balance is that it be large enough to cover the checks written against the balance.

82. Transaction Costs

Transaction or **contracting costs** represent the explicit or implicit costs of facilitating exchanges. For example, firms cannot costlessly issue debt and repurchased equity, or negotiate bank loans. Loan covenants may restrict management's discretion in some decision, or even limit returns to shareholders. Covenants also may increase firm expenses by requiring audits or the periodic review of financial statements by the lenders. Real-world firms must pay several different categories of transaction costs such as **flotation costs**, **bankruptcy costs**, **agency costs**, and **information asymmetry**. [See also **Flotation costs**, **Bankruptcy costs**, **Agency costs**, and **Information asymmetry**]

83. Transactions Account

Deposit account on which a customer can write checks. Those account include demand deposit and NOWs accounts, NOW represents negotiable order of withdrawal.

84. Transactions Motive

A reason for holding cash that arises from normal disbursement and collection activities of the firm. [See **Transaction cash**]

85. Transfer Pricing (Financial Institution)

The pricing of funds transferred between organizational units of a bank, such as determining the cost of collecting deposits and borrowed funds to finance a loan.

86. Transfer Pricing (Manufacture Firm)

It refers to the divisional income determination for deriving the appropriate price at which goods and services should be transferred from one organizational segment to another. The transfer price represents a sales price to the selling segment and a cost price to the buying segment. The transfer price, therefore, significantly affects reported profits of both segments and divisions.

Transfer prices needed for financial reporting purposes may differ from those required for internal decision and management purposes. A particular transfer price base may be excellent for internal performance measurement purposes, for motivating divisional managers, for instituting and maintaining cost control programs, for achieving full utilization of excess capacity, or for the proper allocation of firm resources. However, this same base may be inappropriate for external reporting purposes.

87. Transit Item

Checks drawn on banks outside the community of the bank in which they are deposited. Transit checks deposited are defined as checks drawn on any bank other than the subject bank.

88. Transition Matrix

A square table of probabilities which summarize the likelihood that a credit will migrate from its current credit rating today to any possible credit rating – or perhaps default – in one period.

89. Translog Production Function

In a translog production function, the number of parameters practically “explodes” as the number of considered production factors increases. Consequently, the shortcoming in the estimation of the respective production function is the occurrence of collinearity.

90. Treasurer

The firm’s treasurer oversees the traditional functions of financial analysis: capital budgeting, short-term and long-term financing decisions, and current asset management and usually reports to the chief financial officer (CFO).

91. Treasury Bills

Treasury bills are short-term debt securities issued by the US government. They are perceived to be virtually risk-free; they have essentially no default risk, since investors fully expect the government to pay all interest and principal when it comes due, and their short duration prevents risk due to market movements over time.

T-bills are the most widely traded and, consequently, the most important money market instruments. The Federal Reserve auctions new issues of T-bills with maturities of 91 or 182 days every Monday. Once a month, the Fed offers T-bills with 365-day maturities, as well. Denominations range from \$10,000 to \$100,000 per bill. All obligations for repayment rest with the US government.

92. Treasury Bill Futures

A futures contract on a Treasury bill. The Treasury bill futures contract promise the future delivery of Treasury bill. These contracts were started in 1972. These contracts are one of the most important future contracts used by the financial institutions to hedge interest rate risk.

93. Treasury Inflation Protected Securities (TIPS)

On January 29, 1997, the US Treasury auctioned a new inflation-indexed security, Treasury Inflation Protected Securities (TIPS). The auction was considered by some to be the biggest news in Treasury debt management since the introduction 20 years ago of the 30-year Treasury security.

Inflation-indexed securities provide a degree of inflation protection for investors and potentially represent cost savings for the US Treasury because it will not have to pay premium for inflation uncertainty. The interest rate paid on these securities (known as the “real rate”) provides investors with a guaranteed semiannual return above inflation.

94. Treasury Bond or Note

Debt obligations of the federal government that make semiannual coupon payments and are sold at or near par value in denominations of \$1,000 or more. They have original maturities of more than 1 year. Treasury notes have initial maturity of 10 years or less and treasury bonds have longer maturity.

95. Treasury Bonds Futures

A futures contract on a Treasury bonds. Financial institutions frequently use this kind of future contract to hedge interest rate risk.

96. Treasury Note

Treasury notes have maturities of less than 10 years. [See also **Treasury bond**]

97. Treasury Note Futures

A futures contract on Treasury notes. It is one of the important future contracts used by financial institution to hedge interest rate risk.

98. Treasury Stock

Shares of stock that have been issued and then repurchased by a firm.

99. Tree

A representation of the evolution of the value of a market variable for the purposes of valuing an option or other derivative. [See **Decision tree**]

100. Trend Analysis

One method that can be used to forecast financial data is known as trend analysis. In trend analysis a regression line is fitted to the financial variable over time. A trend line would be fitted using the method of least squares, and this trend line could be used to forecast next year's sales. The following sales model would be estimated:

$$Sales_n = a_0 + a_1n + \varepsilon_n$$

where:

$Sales_n$ = sale in year n

n = year

ε_n = error term

a_0, a_1 = constant to be estimated

101. Treynor's Measure

Ratio of excess return to beta.

$$\frac{\bar{R}_i - R_f}{\beta_i}$$

where $\bar{R}_i$ = average rates of return for i th security or portfolio; R_f = risk free rate; and β_i = beta coefficient.

102. Triangular Arbitrage

Striking offsetting deals among three markets simultaneously to obtain an arbitrage profit.

103. Trinomial Tree

A tree where there are three branches emanating from each node. It is valuing derivatives. It can be used as an alternative to binomial tree. Under this case, the probabilities is classified into up (P_u), middle (P_m), and down (P_d).

104. Triple-Witching Hour

The four times a year that the S&P 500 futures contract expires at the same time as the S&P 100 index option contract and option contracts on individual stocks. It is called triple-witching hour because of the volatility believed to be associated with the expirations in these three types of contracts.

105. Trough

The transition point between recession and recovery for business cycle. [See **Business cycle**]

106. Trust Department

Trust refers a property interest held by trust refers one party for the benefit of another. Trust departments are responsible for managing the investments of individuals or institutional clients such as a pension fund.

107. Trust Receipt

Many businesses lack the financial strength and reputation to support unsecured borrowing. These firms may be able to meet their needs for funds by using physical assets to secure the loan. In such cases, the lender takes out a trust receipt, that is, a lien, against these assets. Inventory is the asset most commonly used to secure borrowing in this way.

The lender protects itself against risk by advancing only a portion of the estimated market value of the assets. Where the inventory is readily transferable and salable, the lender may advance as much as 90%. If the inventory is highly specialized, however, the proportion is likely to be considerably lower.

Straightforward borrowing by a trust receipt presents a serious disadvantage in that the physical property that secures the loan must be described in detail in the legal documents. This is clearly difficult if various finished goods are being pledged. An alternative to this is a **floating lien**. [See also **Floating lien**]

108. Trustee

All bonds will have **indentures**. [See also **Indentures**] Such agreements are supervised by a trustee who acts on behalf of bondholders to ensure proper execution of the indenture provisions by the corporation. If the issuer violates indenture provisions, it is in default and the trustee must act to protect the bondholders' interests.

109. Truth-in-Lending

A law passed by the US Congress in 1968 that requires covered lenders to disclose fully all the relevant terms of a personal loan to the borrower and to report a standardized loan rate (known as the APR, or annual percentage rate).

1.21 U

1. UBPR

UBPR represents Uniform Bank Performance Report. Traditionally, most banks did similar things for their balance sheets in accordance with UBPR. They accepted deposits and made loans, and many interest rates were regulated. The primary differentiation of performance amongst banks was balance sheet composition.

Today, banks may pursue sharply different strategies. While some emphasize traditional deposit-gathering and lending, others offer services such as trusts, mortgage-banking, insurance, brokerage, and asset management. This is particularly true of bank holding companies that can operate subsidiaries across a wide range of businesses such as banking, brokerage and insurance. At a minimum, peer comparisons should be made across firms with similar strategies. Second, a bank's total assets no longer serve as a meaningful yardstick when banks engage in off-balance sheet activities.

2. Unbundling

[See **Bundling**]

3. Uncommitted Line of Credit

An uncommitted line of credit does not have an up-front fee payment and so the bank is not obliged to lend the firm money. If the bank chooses to lend under the terms of the line of credit, it may do so, but is also may choose not to lend. [See also **Revolving credit agreement**]

4. Underlying Asset

The asset whose price determines the profitability of a derivative. For example, the underlying asset for a purchased call is the asset that the call owner can buy by paying the strike price.

5. Underlying Variable

A variable that the price of an option or other derivative depends on. [See **Black-Scholes option pricing model**]

6. Underpricing

Underpricing represents the difference between the aftermarket stock price and the offering price. Underpricing represents money left on the table, or money the firm could have received had the offer price better approximated the aftermarket value of the stock.

7. Underwrite

Purchase securities from the initial issuer and distribute them to investors. [See **Underwriter**]

8. Underwriter

The investment bank carries, or underwrites, the risk of fluctuating stock prices. Thus, an investment bank is sometimes called an underwriter. Should the market's perception of the issuer change or a macroeconomic event result in a stock market decline, the investment bank carries the risk of loss, or at least the possibility of a smaller than expected spread.

9. Underwriting, Underwriting Syndicate

Underwriters (investment bankers) purchase securities from the issuing company and resell them. Usually a syndicate of investment bankers is organized behind a lead firm. [See **Syndicate**]

10. Undivided Profits

Retained earnings or cumulative net income not paid out as dividends. It can be used as an internal source of funds.

11. Unearned Interest

Interest received prior to completion of the underlying contract.

12. Unemployment Rate

The ratio of the number of people classified as unemployed to the total labor force. The unemployment rate is used to determine whether a country's economy is in boom, recession, or normal.

13. Unexpected Losses

A popular term for the volatility of losses but also used when referring to the *realization* of a large loss which, in retrospect, was unexpected.

14. Unfunded Debt

Short-term debt, such as account payable, is the unfunded debt. Cost of capital of an unfunded debt has a risk-free rate interest of zero.

15. Uniform Limited Offering Registration

Several states offer programs to ease the process of public equity financing for firms within their borders. A firm in a state that has enacted a ULOR (Uniform Limited Offering Registration) law can raise \$1 million by publicly selling shares worth at least \$5. This law creates a fairly standardized, fill-in-the blank registration document to reduce a firm's time and cost in preparing an offering.

16. Unilateral Transfers

Gifts of goods and money made by residents of one nation to residents of another nation, with the net amount recorded in the donor nation's balance-of-payments accounts.

17. Unique Risk

[See **Diversifiable risk**]

18. Unit Banking States

States that prohibit branching banking are called units banking states. Since 1994, most of the states become branch banking states.

19. Unit Benefit Formula

Method used to determine a participant's benefit plan by multiplying years of service by the percentage of salary.

20. Unit Investment Trust

Money invested in a portfolio whose composition is fixed for the life of the fund. Shares in a unit trust are called redeemable trust certificates, and they are sold at a premium above net asset value.

21. Unit of Production Method

The unit production method is one of the accelerated depreciation methods. This method determines the depreciation in accordance with total production hours for the machines and production hours operate each year. For example, the expected useful life of 5,000 h is divided into the depreciable cost (cost-salvage value) to obtain an hourly depreciation rate of \$1.08. If we assume the machine is used 2,000 h the first year, 1,000 h the second year, 900 h the third year, 700 h the fourth year, and 400 h the fifth year, then the annual depreciation is determined as follows:

Year 1	$\$1.08 \times 2,000 =$	\$2,160
Year 2	$\$1.08 \times 1,000 =$	1,080
Year 3	$\$1.08 \times 900 =$	972
Year 4	$\$1.08 \times 700 =$	756
Year 5	$\$1.08 \times 400 =$	432
		\$5,400

22. Unit Volume Variability

Variability in the quantity of output sold can lead to variability in EBIT through variations in sales revenue and total variable costs such as raw material costs and labor costs. The net effect of fluctuating volume leads to fluctuations in EBIT and contributes to business risk. [See **Business risk**]

23. Universal Financial Institution

Universal Financial Institution (FI) is an FI that can engage in a broad range of financial service activities. Financial system in the United States has traditionally been structured along separatist or segmented product lines. Regulatory barriers and restrictions have often inhibited the ability of an FI operating in one area of the financial services industry to expand its product set into other areas. This might be compared with FIs operating in Germany, Switzerland, and the United Kingdom, where a more **universal FI** structure allows individual financial services organizations to offer a far broader range of banking, insurance, securities, and other financial services products. However, the recent merger between Citicorp and Travelers to create Citigroup, the third largest universal bank or financial conglomerate in the world, was a sign that the importance of regulatory barrier in the United States is

receding. Moreover, the passage of the Financial Services Modernization Act of 1999 has accelerated the reduction in the barriers among financial services firms. Indeed, as consolidation in the US and global financial services industry proceeds apace, we are likely to see acceleration in the creation of very large, globally oriented multi-product financial service firms.

24. Universal Life Insurance

A form of life insurance policy in which savings contributed by the policyholder are placed in a money market fund, with the life insurance company making periodic withdrawals to cover the premiums owed on the life insurance policy.

25. Universal Life Policy

An insurance policy that allows for a varying death benefit and premium level over the term of the policy, with an interest rate on the cash value that changes with market interest rates. Universal life was introduced in 1979. It combines the death protection features of term insurance with the opportunity to earn market rates of return on excess premiums. Unlike variable life, with its level premium structure, premiums on universal life policies can be changed. The policyholder can pay as high a “premium” as desired, instructing the insurer to invest the excess over that required for death protection in the insurer’s choice of assets. Later, if the policyholder wishes to pay no premium at all, the insurer can deduct the cost of providing death protection for the year from the cash value accumulated in previous years. With other types of policies, skipping a premium would cause the policy to lapse. Unlike whole or variable life policies, the face amount of guaranteed death protection in a universal life policy can be changed at the policyholder’s option. Also, unlike variable life, the cash value has a minimum guaranteed rate of return.

26. Unseasoned New Issue

Initial public offering (IPO). It is the first public equity issue that is made by a company. All initial public offerings are cash offers because if the firm’s existing shareholders wanted to buy the shares, the firm would not need to sell them publicly.

27. Usury Ceilings

Usury refers to interest charges in excess of that legally allowed for a specific instrument. Besides disclosure and bankruptcy laws, some states restrict the rate of interest that may be charged on certain categories of loans – primary consumer loans, but also some agricultural and small business loans. Usury laws establish rate ceilings that a lender may not exceed, regardless of the lender’s costs. Usury ceilings apply to lenders of all types, not just to depository institutions.

28. Unsystematic Risk

A well-diversified portfolio can reduce the effects of firm- or industry-specific events – such as strikes, technological advances, and entry and exit of competitors – to almost zero. Risk that can be diversified away is known as unsystematic risk or **diversifiable risk**. Information that has negative implications for one firm may contain good news for another firm. In a well-diversified portfolio of firms from different industries, the effects of good news for one firm may effectively cancel out bad news for another firm. The overall impact of such news on the portfolio's returns should approach zero.

29. Up-and-In

A knock-in option for which the barrier exceeds the current price of the underlying asset.

30. Up-and-Out

A knock-out option for which the barrier exceeds the current price of the underlying asset.

31. Up-and-Out-Option

An option that comes into existence when the price of the underlying asset increases to a prespecified level.

32. Uptick

A trade resulting in a positive change in a stock price, or a trade at a constant price following a preceding price increase.

33. Utility Function

Utility is the measure of the welfare or satisfaction of an investor. The utility function can be defined as $U = f(\bar{R}, \sigma^2)$ when $\bar{R}$ = average rates of return and σ^2 = variance of rate of return.

34. Utility Theory

Utility theory is the foundation for the theory of choice under uncertainty. Following Henderson and Quandt (1980), cardinal and ordinal theories are the two major alternatives used by economists to determine how people and societies choose to allocate scarce resources and to distribute wealth among one another over time. [See also **Cardinal utility** and **Ordinal utility**]

35. Utility Value

The welfare a given investor assigns to an investment with a particular return and risk. [See **Utility function**]

1.22 V

1. VA Loan

A VA loan is the mortgage which is made by banks and insured by the Veterans Administration (VA) which is a federal agency insuring mortgages.

2. Valuation Reserve

Loan-loss reserve reported on the balance sheet; losses can be charged only against this reserve. In the balance sheet it is listed as loan and lease loss allowance.

3. Value Additivity (VA) Principle

In an efficient market the value of the sum of two cash flows is the sum of the values of the individual cash flows. No matter how the payments are divided among claimants, the sum of the values will be the same. Value of bond + value of stock = value of firm.

4. Value at Risk

Value at risk (VaR) is a procedure for estimating the maximum loss associated with a security or portfolio over a specific period of time, associated with a given confidence level. VaR can be used to measure either market risk or credit risk. In a loss distribution, loss can be either expected loss (EL) or unexpected loss (UL). The UL is considered the measure of VAR.

5. Vanilla Option

A standard option or other derivative. For example, ordinary puts and calls are “vanilla” options.

All vanilla options share a few common characteristics: (a) one underlying asset; (b) the effective starting time is present; (c) only the price of the underlying asset at the option's maturity affects the payoff of the option; (d) whether an option is a call or a put is known when sold; and (e) the payoff is always the difference between the underlying asset price and the strike price, and so on. Vanilla options have many limitations resulting from their lack of flexibility. Each kind of exotic options, to some degree, overcomes one particular limitation of vanilla options.

6. Variable Annuities

Annuity contracts in which the insurance company pays a periodic amount linked to the investment performance of an underlying portfolio. Variable annuities are structured so that the investment risk of the underlying asset portfolio is passed through to the recipient, much as shareholders bear the risk of a mutual fund. There are two stages in a variable annuity contract: an **accumulation phase** and a **payout phase**. [See also **Accumulation phase** and **Payout phase**]

7. Variable Cost

A cost that varies directly with volume and is zero when production is zero. For example, if a cost is \$3/unit, and it has 100 units, its total variable cost is \$300. When the number of units becomes 200, the total variable cost is \$600.

8. Variable Life Policy

An insurance policy that provides a fixed death benefit plus a cash value that can be invested in a variety of funds from which the policyholder can choose. First introduced in 1975, variable life policies gained popularity after 1980 as an insurance vehicle providing some protection against inflation. Like whole life policies, variable life policies require level premium payments throughout the policyholder's life, but there are important differences. For example, excess premiums that add cash value earn variable, not fixed, rates return, based on the insurer's yield on assets of the *policyholder's* choice. If the selected assets perform well, cash value and death benefits both increase. If not, the cash value may be zero, so the insured bears the entire investment risk. A minimum death benefit is specified in the policy, although there is no maximum. The actual payment to beneficiaries depends on yields earned on excess premiums.

9. Variable Universal Life

The newest life insurance product is **variable universal life**, introduced in 1985. So named because it combines the investment flexibility of variable life with the death benefit and premium flexibility of universal life, this new type of policy has gained rapid acceptance among purchasers of life insurance. Variable universal life gives policyholders the greatest freedom to adjust death benefits, premium payments, and investment risk/expected return as their cash-flow and death protection needs change. (Some sources also use the name **flexible premium life** for this new policy.)

10. Variable Rate Securities

A floating rate security refers to the applicable market interest rate is tied to some index and changes whenever the index changes. In other words, a variable rate

security is automatic repricing, usually by changing the interest rate at predetermined intervals. For example a variable rate CD.

11. Variance

The historical risk of an asset can be measured by its variability of its net income in relation to its **arithmetic average**. [See also **Arithmetic average**] The variance, σ^2 , from a sample of data of random variable X is computed by summing the squared deviations and dividing by $n - 1$.

$$\sigma^2 = \frac{\sum_{t=1}^n (X_t - \bar{X})^2}{n - 1},$$

where:

X_t = observation t for random variable X

$\bar{X}$ = arithmetic average of X

N = number of observations for X

Squaring the terms can make the variation difficult to interpret. Therefore, analysts often prefer the **standard deviation**, which is simply the square root of the variance. [See also **Standard deviation**]

12. Variance Rate

It represents variance per unit of time. In a generalized Wiener process has two variables, (a) expected drift rate (average drift per unit of time); and (b) variance rate.

13. Variance Reduction Procedures

Procedures for reducing the error in a Monte Carlo simulation.

14. Variation Margin

An extra margin required to bring the balance in a margin account up to the initial margin when there is a margin call.

15. Vega

The change in the price of a derivative due to a change in volatility. Also sometimes called *kappa* or *lambda*. Based upon the call option formula defined in option pricing model. [See **Option pricing model**] The mathematical result can be defined as:

$$\frac{\partial C}{\partial \sigma} = S\sqrt{T}N'(d_1) > 0$$

where $N'(d) = \frac{2N(d_1)}{2\sigma}$

16. Vega-Neutral Portfolio

A portfolio with a Vega of zero.

17. Vehicle Currency

A monetary unit of a nation that is not only the standard of value (unit of account) for domestic transactions, but also used to express the prices of many goods and services traded between other nations as well.

18. Venture Capital

Venture capitalists invest funds in private companies in return for ownership shares. Venture capital comes from a pool of money raised from a variety of limited partners, such as pension funds, insurance companies, and wealthy individuals; the venture capitalists act as the pool's general partners. The venture capitalist generally invests this capital in equity shares of private firms.

Venture capital does not solve the problem of ownership dilution, especially since venture capitalists often demand large ownership shares in exchange for their funds. The arrangement does have advantages, though. Venture capitalists often have expertise in the technology or marketing needs of the firms in which they invest. Venture capitalists frequently sit on their investees' board of directors and offer technical, marketing, and financial advice. Thus, they provide both funds and expertise to the growing firm.

Of course, venture capitalists do not provide their time and money simply as a public service. They invest with a future goal of "cashing out," or selling their shares in the company for much more than they paid. A venture capitalist cashes out if the firm goes public, if it is acquired by another firm, or if the firm's success allows the original owners to repurchase the venture capitalist's shares at a fair price. The venture capitalist returns the investment's profits to the pool's limited partners.

19. Vertical Acquisition

Acquisition in which the acquired firm and the acquiring firm are at different steps in the production process. The acquisition by an airline company of a travel agency would be a vertical acquisition. There are three types of acquisition, which includes horizontal acquisition, vertical acquisition, and conglomerate acquisition.

20. Vertical Combination

A type of business combination that may involve two firms those are in a supplier-customer relationship. [See **Vertical acquisition**]

21. Vertical Spread

The sale of an option at one strike price and purchase of an option of the same type (call or put) at a different strike price, both having the same underlying asset and time to expiration.

22. VES Production Function

The Variable Elasticity of Substitution (VES) production function framework allows the elasticity of factor substitution to interact with the level of economic development. Properties of VES Production Function include (i) VES satisfies the requirements of a neo-classical production function, (ii) VES function includes the fixed co-efficient models, and (iii) VES production function is more general.

23. Vested Benefits

These refer to benefits that employees are entitled to even if they leave the firm before retirement. The employee is given a legal claim on his or her pension rights when he or she becomes vested. This means that even if the employee leaves the firm, he or she is still entitled to receive a pension from the firm on retirement. There are various types of vesting formulas, which determine when an employee becomes vested. Most formulas are based on the employee's length of service. For example, if a firm's pension policy states that an employee can become vested after working for the firm for 9 years, then after 9 years of working for the firm the employee is entitled to receive a pension. From the firm's perspective, the vesting formula may lower the cost of the pension plan, because employees who leave the company before they become vested are not entitled to receive any pension benefits.

24. Video Conferences

Meetings between business executives or other individuals that are conducted over great distances using satellites and other electronic communications devices.

25. Volatile Deposits

Difference between actual outstanding deposits and core deposits; they represent balances with a high probability of being withdrawn. Implicitly, these are a bank's highly rate sensitive deposit that customers withdraw as interest rates vary.

26. Volatile Funds

For example, negotiable CDs, repurchase agreements, and fed funds purchased are quite volatile. Management assumes that most of these funds could be withdrawn or become unavailable on short notice.

27. Volatility

The standard deviation of the continuously compounded return on an asset. This measure is one of the five variables used to determine the value of option. [See **Black-Scholes option pricing model**]

28. Volatility Index (VIX)

This index was created by the CBOE Global Markets (originally known as the Chicago Board Options Exchange). The VIX is a real-time market index that representing market's expectation of 30-day forward-looking volatility. Derived from the price inputs of the S&P 500 index options, it provides a measure of market risk and investors' sentiments. It is also known by other names like "Fear Gauge" or "Fear Index." In general, if the VIX is figure is larger than 30, investors should decide whether to sell their equity investments. The volatility index can, in general, be represented by the implied variance of the S&P 500 index options. Most recently, researchers suggest that the VIX should use range measurements instead of variance measurements.

29. Volatility Matrix

A table showing the variation of implied volatilities with strike price and time to maturity.

30. Volatility Risk

The risk in the value of options portfolios due to unpredictable changes in the volatility of the underlying asset.

31. Volatility Skew

Generally, implied volatility as a function of the strike price. Volatility skew refers to a difference in premiums as reflected in differences in implied volatility. Skew is sometimes used more precisely to refer to a difference in implied volatilities between in-the-money and out-of-the-money options. [See **Constant elasticity volatility model**]

32. Volatility Smile

A volatility skew in which both in-the-money and out-of-the-money options have a higher volatility than at-the-money options (i.e., when you plot implied volatility against the strike price, the curve looks like a smile).

33. Volatility Swap

Swap where the realized volatility during an accrual period is exchanged for a fixed volatility. Both percentage volatilities are applied to a notional principal. The payments of volatility swap depend upon the volatility of stocks (or other assets).

34. Volatility Term Structure

The variation of implied volatility with time to maturity.

35. Volume

The number of transactions in a futures contract made during a specified period of time.

36. Voluntary Restructuring

Management has three basic approaches to voluntary restructuring. **Carve outs** occur when the parent sells a partial interest in a subsidiary through an IPO. The carve out may increase the selling firm's value due to benefits from restructuring the asset composition of the firm. Again, value is enhanced if the manager focuses more on the remaining assets. **Spin-offs** occur when the parent transfers complete ownership of a subsidiary to the existing shareholders. The spin-off allows the shareholders to retain control over a given asset base while allowing management to focus on a smaller segment of the firm's assets. Finally, **sell offs** involve the direct sale of assets to a third party. The selling firm receives cash, which can be used for debt repayment or reinvestment in the remaining assets. Management in this case cannot only refocus on the main line of core business but also now has the wherewithal to finance any necessary changes.

Any of these voluntary approaches may be used by managers of troubled firms in order to fend off the legal complications stemming from bankruptcy.

37. VRMs

Home mortgage loans carrying a loan interest rate that varies during the term of a loan, generally depending on the movement of interest rates in the open market.

1.23 W

1. Waiting Period

Time during which the Securities and Exchange Commission studies a firm's registration statement. During this time the firm may distribute a preliminary prospectus.

2. Warehousing

A warehousing method of financing can reduce the risk of using inventory as collateral to secure the loan. There are two variations of this method: **field warehousing** and **public warehousing**. [See also **Field Warehousing** and **Public warehousing**]

Warehousing, like receivables financing, is a flexible source of short-term credit that automatically grows as the company's working capital needs expand. Also, like receivables financing, its cost is fairly high. Typically, the warehousing company imposes a service charge, usually a fixed minimum plus 1–2% of the funds loaned, plus an interest rate of 8–12% or sometimes more. The fixed costs of warehousing – the minimum service charge plus the cost of providing the field warehouse facilities or moving goods to a public warehouse – make it unsuitable for very small firms; the minimum feasible inventory size probably is about \$100,000.

3. Warrant

A warrant is a financial instrument issued by a corporation that gives the purchaser the right to buy a fixed number of shares at a set price for a specified period. There usually is a secondary market where existing warrants may be traded.

There are two major differences between a warrant and a **publicly traded option**. [See also **Publicly traded option**] First, the warrant normally matures in 3–5 years, whereas the maturity of a publicly traded option is normally less than 9 months. The second difference is that the warrant is an agreement between the corporation and the warrant's buyer. If the warrant's owner decides to exercise the right to purchase stock, the corporation issues new shares and receives the cash from the sales of those shares.

Typically, a warrant accompanies a bond issue, but it is detachable; it can be traded separately from the bond. A warrant is essentially a call option written by the company that issues the stock. Its value is influenced by the same factors that influence the value of a call option.

In this context, the value of a warrant at expiration (V_W) is defined by the following equation:

$$V_W = \text{Max}[0, NP - NX],$$

where P and X are the price of the stock and the exercise price of the option, respectively; and N is the number of shares obtainable with each warrant.

4. Warren and Shelton Model

Warren and Shelton (1971) devised a simultaneous-equation model. The model has four distinct segments corresponding to the sales, investment, financing, and return-to-investment concepts in financial theory. The entire model is a system of 20 equations of a semi-simultaneous nature. The actual solution algorithm is recursive, between and within segments.

5. Wash

Gains equal losses.

6. Weak-Form Efficient Market

Different assumptions about information availability give rise to different types of market efficiency. [See also **Efficient market**]

A weak-form efficient market is a market in which prices reflect all past information, such as information in last year's annual reports, previous earnings announcements, and other past news. Some investors, called *chartists* or *technicians*, examine graphs of past price movements, number of shares bought and sold, and other figures to try to predict future price movements. A weak-form efficient market implies that such investors are wasting their time; they cannot earn above-average, risk-adjusted profits by projecting past trends in market variables. Generally, evidence indicates that historical information is not helpful in predicting stock price performance. [See **Technicians**]

7. Wealth Effect (of Saving and Interest Rates)

The relationship between the volume of saving and interest rates which contends that the net wealth position of savers (the balance in their portfolios between debt and financial assets) determines how their desired levels of savings will change as interest rates change.

8. Weather Derivatives

Derivative where the payoff depends on the weather.

9. Weekend Effect

The common recurrent negative average return from Friday to Monday in the stock market.

10. Weighted Average Cost of Capital

The weighted average cost of capital (WACC) represents the firm's minimum required rate of return on its average-risk capital budgeting projects. It is found by multiplying the marginal cost of each capital structure component by its appropriate weight and then summing the terms:

$$\text{WACC} = w_d k_d + w_p k_p + w_e k_e$$

The weights of debt, preferred equity, and common equity in the firm's capital structure are given by w_d , w_p , and w_e , respectively. The cost of debt, preferred equity, and common equity are K_d , K_p , and K_e , respectively. As the weighted average cost of capital covers all of the firm's capital financing sources, the weights must sum to 1.0. The firm's cost of common equity, k_e , can reflect the cost of retained earnings, k_{re} , or the cost of new common stock, k_{cs} , whichever is appropriate.

The weights represent a specific, intended mix of debt and equity that the firm will try to achieve or maintain over the planning horizon. As much as possible, the target weights should reflect the combination of debt and equity that management feels will minimize the firm's weighted average cost of capital. It is necessary to minimize the WACC in order to maximize shareholder wealth.

The firm should make an effort over time to move toward and maintain its target capital structure mix of debt and equity.

There are two ways to measure the mix of debt and equity in the firms' capital structure.

One method uses the firms' book values, or balance sheet amounts, of debt and equity. The actual weight of debt in the firm's capital structure equals the book value of its debt divided by the book value of its assets. Similarly, the actual equity weight is the book value of its stockholders' equity divided by total assets. Once the target weights have been determined, the firm can issue or repurchase appropriate quantities of debt and equity to move the balance sheet numbers toward the target weights.

A second method uses the market values of the firm's debt and equity to compare target and actual weights. The actual weight of debt in the firm's capital structure equals the market value of its debt divided by the market value of its assets. Similarly, the actual equity weight is the market value of the firm's stockholders' equity divided by the market value of its assets. Calculated in this way, bond and stock market price fluctuations, as well as new issues and security repurchases, can move the firm toward – or away from – its target.

Financial theory favors the second method as most appropriate. Current market values are used to compute the various costs of financing, so it stands to reason that market-based costs should be weighted by market-based weights.

The basic capital structure of a firm may include debt, preferred equity, and common equity. In practice, calculating the cost of these components is sometimes complicated by the existence of hybrid financing structures (e.g., convertible debt) and other variations of straight debt, preferred equity, or common equity. A

comparison of capital costs between countries also is difficult. What may appear to be lower financing costs in one country may disappear after careful analysis.

11. Weighted Average Life for Mortgage-Backed Securities

The weighted average life (WAL) is a product of the time when principal payments are received and the amount of principal received divided by total principal outstanding. Explicitly, it can be defined as:

$$\text{WAL} = \frac{\sum \text{Time} \times \text{Expected principal received}}{\text{Total principal outstanding}}$$

For example, consider a loan with 2 years to maturity and \$100 million in principal. Investors expect \$50 million of the principal to be repaid at the end of year 1 and the remaining \$60 million to be repaid at maturity.

Time	Expected principal payments	Time $\times$ principal
1	\$50	\$50
2	60	120
	\$100	\$170

$$\text{WAL} = \frac{170}{100} = 1.7 \text{ years}$$

The WAL is presented in mortgage-backed security certificate. In addition to WAL, it also presents: (a) type of security, (b) current price, (c) price change, (d) spread to average life, (e) spread change, (f) prepaid speed, and (g) year to maturity.

12. Weighted Cost of Funds

Weighted average cost of all sources of fund in a depository, including deposits, non-deposits, liabilities, and capital.

13. Weighted Marginal Cost of Fund

Marginal cost of pooled debt funds used in pricing decisions.

14. Weighted Unbiased Estimator

When consideration is given to the types of applications for average rates of return to (1) determine the historical profit rate of an investment and (2) to assess the long-run expected rate of return of some investment instruments, the importance of accuracy and a lack of bias are apparent. Blume (1974) has investigated the possible bias in using either arithmetic average ($\bar{x}$) or geometric average ($\bar{g}$) to forecast such expected rates of return and has proposed four alternative unbiased estimators: (1) simple unbiased, (2) overlapped unbiased, (3) weighed unbiased, and (4) adjusted

unbiased. Blume has also mathematically and empirically shown that the weighted unbiased estimator is the most efficient estimator and is the most robust for non-normal and nonstationary data.

The definition of the weighted unbiased estimator, $M(W)$, is

$$M(W) = \left(\frac{T-n}{T-1}\right)\bar{X} + \left(\frac{n-1}{T-1}\right)\bar{g}$$

where:

T = the number of HPRs used to estimate the historical average returns

n = the number of investment-horizon periods for which a particular investment is to be held

15. Well-Diversified Portfolio

A portfolio spread out over many securities in such a way that the weight in any security is close to zero.

16. White Knight

White knights are alternative suitors (acquirers) that offer friendlier terms to a target firm facing a hostile takeover. [See also **Tender offer**]

17. Whole-Life Insurance Policy

Provides a death benefit and a kind of savings plan that builds up cash value for possible future withdrawal. A whole life policyholder pays pay fixed amount of premiums in exchange for a known death benefit, the face amount of the policy.

18. Wiener Process

A stochastic process where the change in a variable during each short period of time of length δt has a normal distribution with a mean equal to zero and a variance equal to δt . [See also **Brownian motion**]

19. Wild Card Play

The right to deliver on a futures contract at the closing price for a period of time after the close of trading.

20. Wilshire 5000 Equity Index

The Wilshire 5000 equity index, which includes about 7,000 stocks, is compiled by both market-value-weighted and equally weighted approaches. This index is being used increasingly because it contains most equity securities available for

investment, including all NYSE and AMEX issues and the most active stocks traded on the over-the-counter (OTC) market.

The following formula is used to compute the market-value-weighted Wilshire 5000 equity index.

$$I_t = I_{t-1} \left[\sum_{j=1}^N (S_{jt}) P_{jt} / \sum_{j=1}^N (S_{jt-1}) P_{jt-1} \right],$$

where:

I_t = index value for the t th period

N = number of stocks in the index

P_{jt} = price of the j th security for the t th period

S_{jt} = shares outstanding of the j th security for the t th period

P_{jt-1} = price of the j th security for the $(t - 1)$ th period

S_{jt-1} = shares outstanding of the j th security for the $(t-1)$ th period

21. Window Dressing

The practice in financial reporting in which a firm engages in certain transactions at the end of a reporting period (quarter or fiscal year) to make the financial results appear better or different from that prevailing at the time.

22. Winner's Curse

The average investor wins – that is, gets the desired allocation of a new issue – because those who knew better avoided the issue. Winner's curse is the reason why IPOs have a large average return. To counteract winner's curse and attract the average investor, underwriters underprice issues.

23. Witching Day

Witching day is when derivatives traders often use the ordinary stock market to hedge their positions. As a result, as those traders work to close out their positions immediately prior to expiration, they can create a ripple effect on the stock market that in turn can create greater volatility than normal on the ordinary stock exchanges. The takeaway for individual investors is that with institutions having a greater-than-normal presence in the markets on those days, you need to be particularly careful about the impact that professional traders might have on share prices. Anytime you go up against institutions, there is a big risk that you will lose out if you are not careful. Your best move is to maintain your discipline in buying or selling stocks on witching days, or simply to avoid them entirely and choose less frantic sessions to make investment decisions.

24. Whipsawing

Whipsawing occurs when the underlying asset increases enough to trigger rebalancing. After more shares are added, the underlying asset decreases in value

and the additional shares are sold at a lower price than what was paid for them. A common remedy for this problem is to use a larger adjustment gap or filter rule; however, the wrong number of shares would be held if the filter rule were increased, particularly if the stock moved in a linear manner. Whipsawed positions commonly occur when the asset fluctuates around a constant level.

25. Wire Transfers

Wire transfers involve electronic bank-to-bank transfers of funds. A wire transfer can move a large cash balance and make it available to a firm's central finance managers within an hour. While wire transfer is the fastest method available to move funds, it also is the most costly.

26. Working Capital

Working capital is the dollar amount of an organization's current assets, which include cash, marketable securities, accounts receivable, and inventory. These current assets are considered liquid because they can be converted into cash relatively quickly. Each component of working capital is affected by the activities of various parts of the organization. Production, pricing, distribution, marketing, wage contracts, and financing decisions are just a few of the diverse activities within the firm that can affect not only the amount of working capital, but also how quickly the individual assets can be converted into cash.

For example, if the firm's union contract requires that the workers be paid weekly, the amount of cash needed to meet the payroll must be available on each payday. This could require the firm to borrow more cash than if the firm paid its workers only once a month.

The external environment in which the firm operates (product markets, investment markets, and financial markets) also can affect the amount and the rate of change of a firm's working capital. In a highly seasonal industry, inventory typically increases dramatically as demand for the product increases. Inventory then decreases as accounts receivable increases and the inventory is shipped. The cycle is completed when the firm collects cash for its accounts receivable. Many organizations must manage their working capital in the face of seasonal and cyclical forces, which can cause a high degree of variability. [See also **Permanent working capital** and **Temporary working capital**]

27. Working Group

The working group gathers the individuals and firms involved in taking the firm public, including investment banks, law firms, and accounting firms. The firm's management team provides the working group with the necessary information and makes the decisions regarding the public offering process. The members of the group work individually and jointly in a number of areas to try to ensure a successful IPO.

The initial planning for an IPO basically involves getting the firm's legal, financial, and organizational details in proper form to minimize the probability

of difficulties arising either during or after the IPO. The firm will hire auditors to review its past financial statements and past and current accounting practices. The auditors may require changes in accounting methods and a restatement of past financial data to bring them into regulatory compliance under SEC guidelines.

28. Working Hedge Strategy

The Working hedge strategy, formulated by Holbrook Working (1953), makes explicit the speculative aspect of hedging. That is, in any hedged position the basis will not be constant over time. Therefore, the hedger in a certain sense is speculating on the future course of the basis. Yet it is expected that the changes in the basis will involve a smaller degree of risk than the corresponding price risk of the unhedged position.

29. Workout Period

Realignment period of a temporary misaligned yield relationship.

30. World Investable Wealth

The part of world wealth that is traded and is therefore accessible to investors.

31. Writing a Call

Selling a call option.

32. Writing an Option

Selling an option.

33. Written Call

A call that has been sold; a short call.

34. Written Put

A put that has been sold; a short put.

35. Written Straddle

The simultaneous sale of a call and sale of a put, with the same strike price and time to expiration.

1.24 X

1. X Efficiency

X efficiencies are those cost savings not directly due to economies of scope or economies of scale. As such, they are usually attributed to superior management skills and other difficult-to-measure managerial factors. To date, the explicit identification of what composes these efficiencies remains to be established in the empirical banking literature.

1.25 Y

1. Yankee Bonds

US firms are not the only issuers of securities outside their national borders. For example, foreign firms can issue securities in the United States if they follow US security registration procedures. Yankee bonds are US dollar-denominated bonds that are issued in the United States by a non-US issuer. Some issuers also find the longer maturities of Yankees attractive to meet long-term financing needs. While Eurodollar bonds typically mature in 10 years or less, Yankees have maturities as long as 30 years. [See also **Eurodollar bonds**]

2. Year End Effect

[See also **January effect**]

3. Yield

A return provided by an instrument. For example, yield for investing in stock is equal to dividend yield plus capital gain yield.

4. Yield Curve

Diagram relating market interest rates to term-to-maturity on securities that differ only in terms of maturity. Alternatively it implies the set of yields to maturity for bonds with different times of maturity. [See **Stem structure of interest rates**]

5. Yield-Giveup Swap

The yield-giveup swap version of the intermarket-spread swap works against the investor over time. Therefore, when a swap involves a loss in yield, there is a high premium to be placed on achieving a favorable spread change within a relatively short workout period.

6. Yield-Pickup Swap

In a pure yield-pickup swap there is no expectation of market changes, but a simple attempt to increase yield. Basically two bonds are examined to establish their difference in yield to maturity, with a further adjustment to consider the impact of interim reinvestment of coupons at an assumed rate of return between now and the maturity date.

7. Yield Curve Swap

A subset of the basis swap; involves exchange of interest payments indexed to a short-term rate for payments indexed to a long-term.

8. Yield Rate

Tax-equivalent interest income divided by earning assets.

9. Yield to Maturity

The yield to maturity or **market interest rate** is the effective annual rate of return demanded by investors on bonds of a given maturity and risk. To properly discount the semi-annual coupons, we must determine the periodic interest rate that corresponds to the effective annual rate. We can calculate the effective annual rate as:

$$\text{EAR} = \text{YTM} = (1 + \text{Periodic interest rate})^m - 1$$

in order to solve for the periodic interest rate:

$$\text{Periodic interest rate} = (1 + \text{YTM})^{1/m} - 1$$

There are two alternatives to calculate yield to maturity.

Alternative 1: Formal Method

Bond price quotes are available in the marketplace, either from bond dealers or from the daily price listings found in secondary sources, such as *The Wall Street Journal*. Both investors and financial managers must calculate the yield to maturity on bonds, given known par values, coupon rates, times to maturity, and current prices. The yield to maturity can be determined from the present value of an annuity factor (PVIFA) and present value interest factor (PVIF) formulas we used to compute bond price:

$$\text{Price} = \text{PVIFA}(\text{Coupon}) + \text{PVIF}(\text{Par Value})$$

$$\text{Price} = \$CF/2 \left[\frac{1 - \left(\frac{1}{1+r}\right)^n}{r} \right] + \text{Par} \left[\frac{1}{(1+r)^n} \right]$$

where r represents the periodic interest rate; and n is the number of semiannual periods until the bond matures. The yield to maturity equals $(1 + r)^2 - 1$; the stated annual rate equals $r \times 2$. But mathematics offers no simple technique for computing r . It is easier to use available technology to solve for the periodic rate. Financial calculators can generally be used to calculate the YTM. The YTM for a zero coupon bond can be defined as $YTM = \left(\frac{\text{Par}}{\text{Price}}\right)^{\frac{1}{n}} - 1$.

Alternative 2: Approximate Method

For a quick estimate of return, the approximation method can be used. Here, the average annual dollar return to the investor of a bond that matures in n years is the coupon payment plus a straight-line amortization of the bond's premium (or discount):

$$\text{Average annual dollar return} = \text{Annual coupon} + \frac{\text{Par} - \text{Price}}{n}.$$

The average amount invested in the bond is the average of its purchase price and par value:

$$\text{Average investment} = \frac{\text{Par} + \text{Price}}{2}.$$

The approximate yield to maturity

$$= \frac{\text{Annual Coupon} + \frac{\text{Par} - \text{Price}}{n}}{\frac{\text{Par} + \text{Price}}{2}}$$

1.26 Z

1. Z-Score Model

Z-score is a statistical measure that presumably indicates the probability of bankruptcy. [See **Credit scoring model**]

2. Z-Tranche

The final class of securities in a CMO exhibiting the longest maturity and greatest price volatility. These securities often accrue interest until all other classes are retired.

3. Zero-Balance Accounts

Zero-balance accounts (ZBAs) centralize cash control at the main corporate office. The zero-balance account is a specialized disbursement account on which

the firm writes checks against a zero balance. Authorized employees write checks on their departmental accounts, but the firm maintains no funds in these accounts. Instead, these accounts accumulate negative bank balances daily. The cash-control system corrects these daily negative balances by releasing funds from a corporate master account, restoring them to zero balances each day.

A zero-balance account offers a firm with many operating divisions several benefits:

- Greater centralized control over disbursements
- Elimination of redundant idle bank balances in different banks
- Reduction of cash transfer expenses
- More effective cash investments
- Greater autonomy for local managers

A ZBA system does require the firm to maintain all accounts at the same concentration bank, however.

4. Zero-Beta Portfolio

The minimum-variance portfolio uncorrelated with a chosen efficient portfolio. This portfolio has beta equal to zero.

5. Zero-Cost Collar

The purchase of a put and sale of a call where the strikes are chosen so that the premiums of the two options are the same.

6. Zero Coupon Bonds

Zero coupon bonds pay no coupon interest and provide only one cash flow: payment of their par value upon maturity. Treasury bills are a form of zero coupon debt. An investor purchases a T-bill at a price below par and receives no interest or other cash flows until maturity. At that time, the investor receives the par value of the T-bill. The return on the security is the difference between its discount price and its par value.

The primary reason for the popularity of zero coupon bonds is that investors do not face any reinvestment rate risk. As these bonds provide no cash flows to reinvest, investors effectively lock in a given yield to maturity. However, under IRS regulations, investors must pay yearly taxes on the *implicit* interest paid by the bonds; the IRS has special rules for determining this value. In essence, investors must pay taxes on income they have not received. Thus, zero coupon bonds are mainly purchased by tax-exempt investors who pay no tax on their investment returns, such as pension funds.

Issuing a zero coupon bond also helps to lower borrowing costs for the firm. The original discount can be expensed for tax purposes on a straight-line basis over the

life of the bond. Thus, rather than cash outflows from coupon interest payments, the issuing firm receives annual cash inflows from tax savings. However, the issuer must plan for a large capital requirement at the maturity of these bonds.

7. Zero-Coupon Interest Rate

The interest rate that would be earned on a bond that provides no coupons.

8. Zero-Coupon Swap

All cash flows of the swap occur at the end of the life of the agreement; payment obligations are compounded to future maturity.

9. Zero-Coupon Yield Curve

The set of yields to maturity for zero-coupon bonds with different times to maturity. [See also **Yield to maturity** for a discussion of calculations]

10. Zero Gap

Gap can be either positive, negative or zero. Zero gap implies that rate sensitive asset equal rate sensitive liability.

11. Zero-Investment Portfolio

A portfolio of zero net value, established by buying and shorting component securities, usually in the context of an arbitrage strategy.

12. Zero-Plus Tick

[See **Uptick**]

13. Zero Rate

[See **Zero-coupon interest rate**]

References

- Altman, E.I. 1968. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance* 23: 589–609.
- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–659.
- Brotherton-Ratcliffe. 1994. Monte Carlo monitoring. *Risk*: 53–58.
- Cecchetti, S.G., R.E. Cumby, and S. Figlewski. 1988. Estimation of the optimal futures hedge. *Review of Economics and Statistics* 70: 623–630.

- Chen, S.S., C.F. Lee, and K. Shrestha. 2001. On a mean-generalized semivariance approach to determining the hedge ratio. *Journal of Futures Markets* 21: 581–598.
- Derman, E., D. Ergener, and I. Kani. 1995. Static options replication. *The Journal of Derivatives* 2 (4): 78–95.
- Elgers, P. 1980. Accounting-based risk predictions: A re-examination. *The Accounting Review* 55 (3): 389–408.
- Fama, E.F., and J.D. MacBeth. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 81: 607–636.
- Fama, E.F., E.L. Fisher, M. Jensen, and R. Roll. 1969. The adjustment of stock prices to new information. *International Economic Review* 10: 1–21.
- Henderson, J.M., and R.E. Quandt. 1980. *Micro economic theory: A mathematical approach*. 3rd ed, 363–391. McGraw-Hill, New York. Appendix.
- Hopewell, M.H., and G.G. Kaufman. 1973. Bond price volatility and term to maturity: A generalized respecification. *American Economic Review*: 749–753.
- Howard, C.T., and L.J. D'Anton. 1984. A risk-return measure of hedging effectiveness. *Journal of Finance and Quantitative Analysis* 1984: 101–112.
- Hung, J.C., C.L. Chiu, and M.C. Lee. 2006. Hedging with zero-value at risk hedge ratio. *Applied Financial Economics* 16: 259–269.
- Johnson, L.L. 1960. The theory of hedging and speculation in commodity futures. *Review of Economic Studies* 27: 139–151.
- Larson, H., and N. Gonedes. 1969. Business combinations: An exchange-ratio determination model. *Accounting Review*: 720–728.
- Latane, H.A., and R.J. Rendleman. 1976. Standard deviation of stock price ratios implied by option Premia. *Journal of Finance* 31: 369–382.
- Lee, C.F., P. Newbold, J.E. Finnerty, and C.C. Chu. 1986. On accounting-based, market-based, and composite-based beta predictions: Methods and implications. *Financial Review*: 51–68.
- Lee, C.F., A.C. Lee, J. Lee, and M. Lee. 2022. *Corporate finance and strategy: An active learning approach*. 2nd ed. World Scientific.
- Lev, B. 1969. Industry averages as targets for financial ratios. *Journal of Accounting Research* 7: 290–299.
- Li, D.X. 2000. *On default correlation: A copula function approach*. Risk Metrics Group.
- Lien, D., and Y.K. Tse. 1998. Hedging time-varying downside risk. *Journal of Futures Markets* 18: 705–722.
- Lintner, J. 1956. Distribution and incomes of corporations among dividends, retained earnings and taxes. *American Economic Review*: 97–113.
- Markowitz, H.M. 1952. Portfolio selection. *Journal of Finance* 7: 77–91.
- Markowitz, H.M. 1959. Portfolio Selection. Cowles Foundation Monograph 16. New York: John Wiley and Sons, Inc.
- Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica* 4 (5): 867–887.
- Miller, M., and F. Modigliani. 1961. Dividend policy, growth and the valuation of shares. *Journal of Business* 43: 411–432.
- Rendelman, R.J., and B.J. Bartter. 1979. Two-stat option pricing. *Journal of Finance* 34.
- Ross, S.A. 1976. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* 13: 341–360.
- Rubinstein, M. 1985. Alternative paths for portfolio insurance. *Financial Analyst Journal* 41 (July/August): 42–52.
- Rubinstein, M., and H.E. Leland. 1981. Replacing options with positions in stock and cash. *Financial Analysts Journal* 37 (July/August): 63–72.
- Saunders, A., and M.M. Cornett. 2011. *Financial institutions management: a risk management approach*. 7th ed. McGraw-Hill.
- Schroder, M. 1989. Computing the constant elasticity of variance option pricing formula. *Journal of Finance* 44 (1): 211–218.
- Shalit, H. 1995. Mean-Gini hedging in futures markets. *Journal of Futures Markets* 15: 617–635.

- Sharpe, W.F. 1964. Capital assets prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- . 1967. A linear programming algorithm for mutual fund portfolio selection. *Management Science* March: 499–510.
- . 1984. Factor models, CAPM's and the APT. *Journal of Portfolio Management* 11: 21–25.
- Treynor, J., and F. Black. 1973. How to use security analysis to improve portfolio selection. *Journal of Business* 46: 66–86.
- Vasicek Oldrich, A. 1973. A note on using cross-sectional information in Bayesian estimation of security betas. *Journal of Finance* 28: 1233–1239.
- Warren, J. and J. Shelton. "A simultaneous equations approach to financial planning." *Journal of Finance*, 26 (September 1971, pp. 1123-1142).
- Working, H. "Hedging reconsidered" *Journal of Farm Economics*, Volume 35 (June 1953, pp. 544-561).

Part II

Papers



Deposit Insurance Schemes

2

James R. Barth, Nguyen Nguyen, and Jiayi Xu

Abstract

More than two-thirds of member countries of the International Monetary Fund have experienced one or more banking crises in recent years. The inherent fragility of banks has motivated more than 70% of the countries in the world to establish deposit insurance schemes. By increasing depositor confidence, deposit insurance has the potential to provide for a more stable banking system. Although deposit insurance increases depositor confidence, however, it removes depositor discipline. Banks may therefore engage in activities that are riskier than would otherwise be the case. Deposit insurance itself, in other words, could increase the likelihood of a crisis. Our purpose is to examine the extent of deposit schemes in countries and the types of schemes adopted as well as assess the benefits and costs of the schemes in promoting stability in the banking sector.

Keywords

Banks · Banking crisis · Banking instability · Bank runs · Bank supervision · Depositor discipline · Deposit insurance · Financial systems · Moral hazard · Regulation

J. R. Barth (✉)

Auburn University and Milken Institute, Auburn, AL, USA

Department of Finance, Harbert College of Business, Auburn University, Auburn, AL, USA

e-mail: barthjr@auburn.edu

N. Nguyen

Minnesota State University, Mankato, MN, USA

e-mail: Nguyen.Nguyen.2@MNSU.edu

J. Xu

Mount St. Joseph University, Cincinnati, OH, USA

e-mail: jzx0022@auburn.edu

2.1 Introduction

Over the past half-century, more than two-thirds of member countries of the International Monetary Fund (IMF) have experienced one or more banking crises¹. These crises occurred in countries at all levels of income and in all parts of the world. This troublesome situation demonstrates that while banks are important for channeling savings to productive investment projects, they nonetheless remain relatively fragile institutions. When a country's banking system experiences systemic difficulties, moreover, the results can be quite disruptive and costly for an entire economy. Indeed, the banking crises that struck many Southeast Asian countries in mid-1997 cost Indonesia alone more than 50% of its gross domestic product (GDP). The most recent global financial crisis of 2007–2009 was even more costly insofar as it involved far more countries, and more industrialized countries.

The inherent fragility of banks has motivated many nations to establish deposit insurance schemes. The purpose of such schemes is to assure depositors that their funds are safe by having the government guarantee that the deposits can always be returned without loss. To the extent that depositors believe that the government will be willing and able to keep its promise, they will have no incentive to engage in widespread bank runs to withdraw their funds. By increasing depositor confidence sufficiently to prevent such runs, deposit insurance thus has the potential to provide for a more stable banking system.

Although deposit insurance increases depositor confidence, however, it gives rise to what is referred to as “moral hazard” (Gropp and Vesala 2004; Gropp et al. 2014; Anginer et al. 2014; Ngalawa et al. 2016). This is a potentially serious problem because depositors who believe their funds are safe have little, if any, incentive to monitor and influence the activities of banks. When this type of depositor discipline is removed, banks are freer to engage in activities that are riskier than would otherwise be the case. To the extent that this type of moral hazard is not kept in check by the bank regulatory and supervisory authorities after a country establishes a deposit insurance scheme, its banking system may still be susceptible to a crisis. Deposit insurance itself, in other words, could contribute to a crisis (Cooper and Ross 2002).

The establishment of a deposit insurance scheme, therefore, is not a sinecure. It provides both potential benefits and costs to society. The difficult issue is determining the optimal trade-off between the benefits and the costs. It is for this reason that governments and citizens in countries around the globe need a better appreciation and understanding of the role of deposit insurance. This is particularly the case insofar as ever more countries have been establishing such schemes in recent years. Indeed, since the first national deposit insurance scheme was established by the United States in 1933 (Bradley 2000), more than 130 countries have done so, many within the past 20 years. The IMF, moreover, suggests that every country should establish one (Garcia 2000).

¹This chapter is based on an earlier version that was published and coauthored with Cindy Lee and Triphon Phumiwasana in 2013.

2.2 The Inherent Fragility of Banks

It is a well-known and widely accepted fact that banks are an important part of a nation's financial system. They complement the nonbank financial institutions and the capital markets in promoting economic growth and development. In particular, banks extend credit to business firms for various investment projects and otherwise assist them in coping with various types of financial risk. They also facilitate the payment for goods and services by providing a medium of exchange in the form of demand deposits. But in providing these services, banks create longer-term assets (credit) funded with shorter-term liabilities (deposits). Therein lies the inherent source of bank fragility. Depositors seeking liquidity may decide to withdraw their deposits from banks at any time. In response, banks may have to sell illiquid assets on short notice and incur losses in the process. Widespread withdrawals may turn a liquidity problem into an insolvency problem.

The worst-case scenario is one in which depositors nationwide become so nervous about the safety of their deposits that they simultaneously decide to withdraw their deposits from the entire banking system. Such a systemic run would force banks to liquidate their assets to meet the withdrawals. A massive sale of relatively opaque assets, in turn, would require that they be sold at "fire-sale" prices to obtain the needed cash. This situation could force illiquid but otherwise solvent institutions into insolvency.

The typical structure of a bank's balance sheet is therefore necessarily fragile. Any bank would be driven into insolvency if its assets had to be immediately sold to meet massive withdrawals by its depositors. This would not be a concern if such an event were a mere theoretical curiosity. However, there has been widespread bank runs in various countries at various points in time. There have even been instances where bank runs in one country have spread beyond its borders to banks in other countries. Unfortunately, bank runs are not benign. They can be destructive insofar as they disrupt both the credit system and the payments mechanism in a country. Worse yet, the bigger the role banks play in the overall financial system of a country, the more destructive a banking crisis can be on economic and social welfare. Unfortunately, this is more likely to be the situation in developing countries.

2.3 The Benefits of Deposit Insurance Schemes

The primary purpose of a deposit insurance scheme is to decrease, if not eliminate, the likelihood of bank runs. A secondary purpose is to protect small depositors from losses. At the time of the Great Depression in the United States, banks had occasionally experienced widespread runs and suffered substantial losses on asset sales in an attempt to meet deposit withdrawals. The situation was so devastating for banks that President Roosevelt declared a bank holiday. When banks were reopened, they did so with their deposits insured by the federal government. This enabled depositors to be confident that their funds were now indeed safe, and therefore there was no need to withdraw them. This action by the government was sufficient to restore

confidence in depositors that their funds were safe in banks. By establishing such a “safety net” for depositors of banks, bank runs were eliminated in the United States.

Before the establishment of deposit insurance in the United States, it was the responsibility of the Federal Reserve System to prevent bank runs. The Federal Reserve was to lend funds to those banks which were experiencing liquidity problems and not solvency problems. In other words, the Federal Reserve was to be a lender of the last resort, always ready to lend to illiquid but solvent banks, when nobody else was willing to do so. Yet, it did not fulfill its responsibility during the 1930s. It was therefore considered necessary to establish an explicit deposit insurance scheme to reassure depositors that their deposits would always be safe and readily available on demand. Deposit insurance thus became the first line of defense against bank runs.

For nearly 50 years after its establishment, the US deposit insurance scheme worked as intended. There were no bank runs and the consensus was that deposit insurance was a tremendous success. But then events occurred that called this view into question. Savings and loans, which had also been provided with their deposit insurance scheme at the same time as banks, were devastated by interest rate problems at first, and then by asset quality problems during the 1980s. The savings and loan problems were so severe that even their deposit insurance fund became insolvent during the mid-1980s. Ultimately, taxpayers were required to contribute the majority of the \$155 billion for cleaning up the mess. Fortunately, even though the deposit insurance fund for banks became insolvent during the late 1980s, the cleanup cost was only about \$40 billion. And taxpayers were not required to contribute to covering this cost. More recently, due to the housing boom and bust and Great Recession almost 500 banks failed from 2008 through 2013 at a cost of approximately \$73 billion to the Deposit Insurance Fund (FDIC [2017](#)).

The fact that several thousand depository institutions – in this case, both savings and loans, and banks – could fail, and cost so much to resolve demonstrated to everyone that deposit insurance was not a panacea for solving banking problems. Despite being capable of addressing the inherent fragility problem of banks, deposit insurance gave rise to another serious problem, namely, moral hazard.

2.4 The Costs of Deposit Insurance Schemes

While instilling confidence in depositors that their funds are always safe, and thereby to prevent bank runs, deposit insurance simultaneously increases the likelihood of another serious banking problem in the form of moral hazard. By removing all concerns that depositors have over the safety of their funds, deposit insurance also removes any incentive depositors have to monitor and influence the activities of banks. Regardless of the riskiness of the assets that are acquired with their deposits, depositors are assured that any associated losses will be borne by the deposit insurance fund, and not by them. This situation, therefore, requires that somebody else must impose discipline on banks. In other words, the bank regulatory and supervisory authorities must now play a bigger role without the help of depositors.

There is widespread agreement that regulation and supervision are even more important to prevent banking problems once countries have established a deposit insurance scheme. Countries doing so must more than ever contain the incentive for banks to engage in excessively risky activities once they have access to deposits insured by the government. The difficult task, however, is to replace the discipline of the private sector with that of the government. Nonetheless, it must and has been done with varying degrees of success in countries around the world. The proper way to do so involves both prudential regulations and effective supervisory practices.

Skilled supervisors and appropriate regulations can help prevent banks from taking on undue risk, and thereby exposing the insurance fund to excessive losses. At the same time, however, banks must not be so tightly regulated and supervised that they are prevented from adapting to a changing financial marketplace. If this happens, banks will be less able to compete and thus more likely to fail. The regulatory and supervisory authorities must therefore strike an appropriate balance between being too lenient and too restrictive, to promote a safe and sound banking industry.

The appropriateness of specific regulations and supervisory practices necessarily depends upon the specific design features of a deposit insurance scheme. Some features may exacerbate moral hazard, whereas others may minimize it. In other words, a government needs to realize that when designing a scheme, one must take into account the effects the various features will have on both depositor confidence and moral hazard. In this regard, information has recently become available describing many of the important differences among deposit insurance schemes that have been established in a large number of countries. It is, therefore, useful to examine this “menu of deposit insurance schemes.” One can thereby appreciate how these schemes differ, and then try to assess which combination of features seems to strike a good balance between instilling depositor confidence to eliminate bank runs and yet containing the resulting moral hazard that arises when depositor discipline is substantially, if not entirely, eliminated.

2.5 Differences in Deposit Insurance Schemes Across Countries

Of the approximately 220 countries in the world, available information indicates that 135 of them have established or plan to establish deposit insurance schemes. Information on the existence of schemes as well as selected design features in 133 countries is presented in Table 1. It is quite clear from this information that there are important differences in key features across all these countries, which includes both emerging market economies and mature economies (Anginer et al. 2019; Demirgüç-Kunt et al. 2015; Demirgüç-Kunt and Kane 2002; Demirgüç-Kunt and Sobaci 2001; Demirgüç-Kunt and Detragiache 2000; Garcia 1999; The World Bank 2020). At the outset, it should be noted that the vast majority of these countries have only recently established deposit insurance for banks. Indeed, 99 of the 135 countries have established their schemes since 1990. And 26 of these countries established them within the past decade. More countries are either in the process or

Table 1 Design features of deposit insurance schemes in countries around the world

Country	Deposit insurance system ⁴	Year deposit insurance first introduced ⁵	Power to cancel or revoke ^{7,4}	Coverage limit (US\$) ⁶	Is there formal coinsurance? ^{7,4}	Risk-adjusted premium? ^{7,4}
Albania	Yes	2002	No	23,012	No	No
Algeria	Yes	2003	N/A	16,794	No	No
Angola	Yes	2015	N/A	19,203 ²	N/A	N/A
Antigua and Barbuda	No	N.A.	N.A.	N.A.	N.A.	N.A.
Argentina	Yes	1979	No	25,130	No	Yes
Armenia	Yes	2005	No	20,846	No	Yes
Aruba	No	N.A.	N.A.	N.A.	N.A.	N.A.
Australia	Yes	2008	No	174,825	No	No
Austria	Yes	1979	Yes	112,208	No	Yes
Azerbaijan	Yes	2007	No	N.A.	No	No
Bahamas	Yes	1999	N/A	50,000	No	No
Bahrain	Yes	1993	No	\$177 ² or 75% of eligible deposits	No	No
Bangladesh	Yes	1984	Yes	1,178	No	Yes
Barbados	Yes	2007	N/A	12,500	No	No
Belarus	Yes	1996	No	Full	No	No
Belgium	Yes	1974	Yes	112,208	No	Yes
Belize	No	N.A.	N.A.	N.A.	N.A.	N.A.
Benin³	Yes	2014	Yes	2,389	No	N/A
Bermuda	Yes	2011	Yes	25,000	No	Yes
Bhutan	No	N.A.	N.A.	N.A.	N.A.	N.A.
Bolivia	Yes	2001	No	N/A	No	No
Bosnia and Herzegovina	Yes	1998	Yes	28,604	No	No
Botswana	No	N.A.	N.A.	N.A.	N.A.	N.A.

Brazil	Yes	1995	No	62,033	No	No
British Virgin Islands	Yes	2016	N/A	100,000	No	No
Brunei	Yes	2011	N/A	37,313	No	No
Bulgaria	Yes	1996	Yes	112,644	No	Yes
Burkina Faso³	Yes	2014	Yes	2,389	No	N/A
Burundi	No	N.A.	N.A.	N.A.	N.A.	N.A.
Cabo Verde	Yes	2016	N/A	10,840 ²	No	No
Canada	Yes	1967	Yes	N/A	No	Yes
Cayman Islands	No	N.A.	N.A.	N.A.	N.A.	N.A.
Central African Economic and Monetary Community	Yes	2011	N/A	8,606	No	No
Chile	Yes	1986	No	N/A	No	No
China	Yes	2015	N/A	76,675 ²	No	No
Colombia	Yes	1985	No	15,258	No	Yes
Comoros	Yes	N/A	Yes	N/A	N/A	N/A
Cook Islands	No	N.A.	N.A.	N.A.	N.A.	N.A.
Costa Rica	No	N.A.	N.A.	N.A.	N.A.	N.A.
Croatia	Yes	1997	No	112,208	Yes	Yes
Curacao	Yes	N/A	No	N/A	No	No
Cyprus	Yes	2000	Yes	112,208	No	Yes
Czech Republic	Yes	1994	No	112,208	No	Yes
Côte d'Ivoire³	Yes	2014	Yes	2,389	No	N/A
Denmark	Yes	1987	N/A	15,009	No	N/A
Djibouti	No	N.A.	N.A.	N.A.	N.A.	N.A.
Dominican Republic	Yes	1962	No	N/A	No	No
Ecuador	Yes	1998	No	32,000	No	Yes

(continued)

Table 1 (continued)

Country	Deposit insurance system ⁴	Year deposit insurance first introduced ⁵	Power to cancel or revoke ^{7,4}	Coverage limit (US\$) ⁶	Is there formal coinsurance? ^{7,4}	Risk-adjusted premium? ^{7,4}
El Salvador	Yes	1999	No	10,227	No	No
Estonia	Yes	1998	No	112,208	No	Yes
Eswatini	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Fiji	No	N.A.	N.A.	N.A.	N.A.	N.A.
Finland	Yes	1969	No	112,208	No	Yes
France	Yes	1980	No	112,208	No	Yes
Gambia, The	No	N.A.	N.A.	N.A.	N.A.	N.A.
Georgia	No	N.A.	N.A.	N.A.	N.A.	N.A.
Germany	Yes	1966	No	112,208 ⁷	No	Yes
Ghana	Yes	2016	N/A	1,130	No	No
Gibraltar	Yes	1998	No	132,626	No	Yes
Greece	Yes	1995	No	112,208	No	Yes
Guatemala	Yes	1999	No	2,640 ¹	No	Yes
Guernsey	Yes	2008	No	68,875 ²	No	No
Guinea-Bissau ³	Yes	2014	Yes	2,389	No	N/A
Guyana	No	N.A.	N.A.	N.A.	N.A.	N.A.
Haiti	No	N.A.	N.A.	N.A.	N.A.	N.A.
Honduras	Yes	1999	No	9,752	No	No
Hong Kong SAR, China	Yes	2004	No	64,103	No	Yes
Hungary	Yes	1993	No	112,041	No	Yes
Iceland	Yes	1985	No	112,208	No	Yes
India	Yes	1961	Yes	7,004	No	No
Indonesia	Yes	1998	No	143,164	No	No
Ireland	Yes	1989	No	112,208	No	Yes

Isle of Man	Yes	1991	N/A	66,313	No	No
Israel	No	N.A.	N.A.	N.A.	N.A.	N.A.
Italy	Yes	1987	No	112,208	No	Yes
Jamaica	Yes	1998	N/A	4,480	No	Yes
Japan	Yes	1971	No	92,081	Yes	No
Jersey	Yes	2009	No	69,168 ²	No	No
Jordan	Yes	2000	No	70,621	No	No
Kazakhstan	Yes	1999	N/A	39,351	No	No
Kenya	Yes	1988	Yes	5,000	Yes	No
Korea, Rep.	Yes	1996	No	43,185	No	Yes
Kosovo	Yes	2011	No	5,610	N/A	Yes
Kuwait	Yes	2008	No	N/A	No	No
Kyrgyz Republic	Yes	2008	No	2,872	No	No
Lao PDR	Yes	1999	N/A	5,627	No	No
Latvia	Yes	1998	No	112,208	No	Yes
Lebanon	Yes	1967	No	3,274 ²	No	No
Lesotho	No	N.A.	N.A.	N.A.	N.A.	N.A.
Liberia	No	N.A.	N.A.	N.A.	N.A.	N.A.
Liechtenstein	Yes	1992	No	103,359	No	No
Lithuania	Yes	1996	Yes	112,208	No	Yes
Luxembourg	Yes	1989	No	112,208	No	Yes
Macao SAR, China	Yes	2012	No	62,332	No	No
Macedonia, FYR	Yes	1996	Yes	33,682	No	No
Madagascar	No	N.A.	N.A.	N.A.	N.A.	N.A.
Malawi	No	N.A.	N.A.	N.A.	N.A.	N.A.
Malaysia	Yes	1998	Yes	61,080	No	Yes

(continued)

Table 1 (continued)

Country	Deposit insurance system ¹	Year deposit insurance first introduced ⁵	Power to cancel or revoke ^{7,4}	Coverage limit (US\$) ⁶	Is there formal coinsurance? ^{7,4}	Risk-adjusted premium? ^{7,4}
Maldives	Yes	2015	No	1,940 ²	No	Yes
Mal³	Yes	2014	Yes	2,389	No	N/A
Malta	Yes	2003	N/A	112,208	Yes	Yes
Marshall Islands	No	N.A.	N.A.	N.A.	N.A.	N.A.
Mauritania	No	N.A.	N.A.	N.A.	N.A.	N.A.
Mauritius	No	N.A.	N.A.	N.A.	N.A.	N.A.
Mexico	Yes	1986	No	135,222 ⁷	No	No
Moldova	Yes	2004	No	12,143 ²	No	No
Montenegro	Yes	2004	No	56,104	No	No
Montserrat	No	N.A.	N.A.	N.A.	N.A.	N.A.
Morocco	Yes	1996	No	8,772	No	No
Mozambique	Yes	2010	No	332	Yes	No
Namibia	No	N.A.	N.A.	N.A.	N.A.	N.A.
Nepal	Yes	2010	Yes	2,495 ²	No	No
Netherlands	Yes	1978	Yes	112,208	No	Yes
New Zealand	No	N.A.	N.A.	N.A.	N.A.	N.A.
Nicaragua	Yes	2001	No	10,000	No	Yes
Niger³	Yes	2014	Yes	2,389	No	N/A
Nigeria	Yes	1988	Yes	1,629	N/A	Yes
Norway	Yes	1961	No	227,783	No	Yes
Oman	Yes	1995	No	52,016	No	No

Pakistan	Yes	2016	N/A	1,613	No	No
Palau	No	N.A.	N.A.	N.A.	N.A.	N.A.
Palestine	Yes	2013	N/A	20,000	No	No
Panama	No	N.A.	N.A.	N.A.	N.A.	N.A.
Papua New Guinea	No	N.A.	N.A.	N.A.	N.A.	N.A.
Paraguay	Yes	2003	No	25,414	No	No
Peru	Yes	1991	No	30,359	No	Yes
Philippines	Yes	1963	Yes	9,875	No	No
Poland	Yes	1995	No	112,134	No	No
Portugal	Yes	1992	No	112,208	No	Yes
Qatar	No	N.A.	N.A.	N.A.	N.A.	N.A.
Romania	Yes	1996	No	112,169	No	Yes
Russian Federation	Yes	2003	No	22,615	No	Yes
Rwanda	Yes	2016	No	535	No	Yes
Samoa	No	N.A.	N.A.	N.A.	N.A.	N.A.
San Marino	Yes	2011	Yes	112,208	No	Yes
Saudi Arabia	Yes	2016	Yes	53,333	No	No
Senegal³	Yes	2014	Yes	2,389	No	N/A
Serbia	Yes	1989	No	56,040	No	No
Seychelles	No	N.A.	N.A.	N.A.	N.A.	N.A.
Singapore	Yes	2006	No	55,671	No	Yes
Slovak Republic	Yes	1996	Yes	112,208	No	Yes
Slovenia	Yes	2001	Yes	112,208	No	Yes
South Africa	No	N.A.	N.A.	N.A.	N.A.	N.A.
Spain	Yes	1977	No	112,208	No	Yes
Sri Lanka	Yes	1987	No	3,303	No	Yes
Suriname	No	N.A.	N.A.	N.A.	N.A.	N.A.

(continued)

Table 1 (continued)

Country	Deposit insurance system ⁴	Year deposit insurance first introduced ⁵	Power to cancel or revoke? ⁴	Coverage limit (US\$) ⁶	Is there formal coinsurance? ⁴	Risk-adjusted premium? ⁴
Sweden	Yes	1996	No	101,963	No	Yes
Switzerland	Yes	1984	No	103,359	No	No
São Tomé and Príncipe	No	N.A.	N.A.	N.A.	N.A.	N.A.
Taiwan, China	Yes	1985	Yes	99,900	No	Yes
Tajikistan	Yes	2004	No	2,994	No	No
Tanzania	Yes	1994	No	655	No	No
Thailand	Yes	1997	No	331,994	No	No
Togo ³	Yes	2014	Yes	2,389	No	N/A
Tonga	No	N.A.	N.A.	N.A.	N.A.	N.A.
Trinidad and Tobago	Yes	1986	No	18,505	No	No
Tunisia	Yes	2018	No	21,440	Yes	No
Turkey	Yes	1983	No	25,253	No	Yes
Turks and Caicos Islands	No	N.A.	N.A.	N.A.	N.A.	N.A.
Uganda	Yes	1994	No	N/A	No	Yes
Ukraine	Yes	1998	No	8,444	No	Yes
United Kingdom	Yes	1982	No	112,732	No	Yes
United States	Yes	1934	Yes	250,000 ⁷	No	Yes
Uruguay	Yes	2002	No	29,252	Yes	Yes
Vanuatu	No	N.A.	N.A.	N.A.	N.A.	N.A.
Vietnam	Yes	1999	No	3,237	No	No

West Bank and Gaza	Yes	2013	No	10,000	No	No
Yemen	Yes	2008	N/A	8,005	No	No
Zimbabwe	Yes	2003	Yes	123	No	No

Notes:

¹In 2006

²Based on exchange rate in April 2021

³The West African Monetary Union Deposit Insurance Fund insures deposits in all eight countries of the West African Monetary Union (Benin, Burkina Faso, Guinea-Bissau, Ivory Coast, Mali, Niger, Senegal and Togo)

⁴Sources: The World Bank's Bank Regulation and Supervision Survey 2019, where the information is not available from this survey, the information is adopted from International Association of Deposit Insurers (IADI) Annual Survey 2020

⁵Sources: Demirguc-Kant et al. (2005), where the information is not available from this source, the information is adopted from International Association of Deposit Insurers (IADI) Annual Survey 2020

⁶Sources: International Association of Deposit Insurers (IADI) Annual Survey 2020, where the information is not available from this survey, the information is adopted from country's deposit insurance regulator's websites

⁷If there are more than one deposit insurer in the country, the coverage reported is the coverage applied for commercial banks

⁸Central African Economic and Monetary Community includes Cameroon, Central African Republic, Chad, Republic of the Congo, Equatorial Guinea, and Gabon

⁹N.A. means Not Applicable, while N/A means Not Available

likely soon to establish a deposit insurance scheme. Differences in each of the other important features noted in the table will now be briefly described in turn.

One key feature of any deposit insurance scheme is the coverage limit for insured depositors. The higher the limit, the more protection is afforded to individual depositors, but the higher the limit, the greater the moral hazard. The limits vary quite widely for countries, ranging from a low of \$123 in Zimbabwe to a high of \$331,994 in Thailand. For purposes of comparison, the limit is \$250,000 in the United States. One problem with these comparisons, however, is that there are wide differences in the level of per capita income among these countries. Ratios that are high multiples of per capita GDP substantially reduce any discipline that depositors might have otherwise imposed on banks.

Apart from coverage limits, countries also differ concerning coinsurance, which may or may not be a part of the deposit insurance scheme. This particular feature, when present, means that depositors are responsible for a percentage of any losses should a bank fail. Only seven countries have such a feature, however. Yet, to the extent that depositors bear a portion of any losses resulting from a bank's failure, they have an incentive to monitor and help control the riskiness of banks. Usually, even when countries adopt coinsurance, the percentage of losses borne by depositors is capped at 10%. Even this relatively small percentage, however, may be enough to attract the attention of depositors when compared to the return they can expect to earn on their deposits, and thereby help to curb moral hazard.

Some countries have elected to establish an *ex-ante*-funded scheme, whereas others have chosen to provide the funds for any losses from bank failures *ex-post*. Of 92 countries, only 15 have chosen to establish an *ex-post* or unfunded scheme. In this case, the funds necessary to resolve bank failures are obtained only after bank failures occur. This type of arrangement may provide a greater incentive for private monitoring and influence because everyone will know that the funds necessary to resolve problems have not yet been collected. And everyone will also know that a way to keep any funds from being collected is to prevent banks from engaging in excessively risky activities. Of course, the degree of monitoring depends importantly on the source of funding. In this regard, there are three alternative arrangements: (1) public funding, (2) private funding, or (3) joint funding. Of these three sources, private funding provides the greatest incentive for private discipline and public funding the least. However, only four of the countries surveyed by the World Bank (2019) fund their deposit insurance schemes solely based on private sources. At the same time, 53 countries rely solely on public funding. 28 of the schemes that are privately funded, moreover, are also either privately or jointly administered.

In addition to the design features already discussed, two other important features must be decided upon when a country establishes a deposit insurance scheme. One is whether in those countries in which premiums are paid by banks for deposit insurance should be risk-based or not (Prescott 2002; Lee et al. 2015). The advantage of risk-based premiums is that they potentially can be used to induce banks to avoid engaging in excessively risky activities. This enables the banking authorities to have an additional tool to contain moral hazard. Yet, in practice, it is extremely difficult to

set and administer such a premium structure. Table 1 shows that slightly less than one-third of the countries have chosen to adopt risk-based premiums.

The last feature to be discussed is the membership structure of a deposit insurance scheme. A country has to decide whether banks may voluntarily join or will be required to join. A voluntary scheme will certainly attract all the weak banks. The healthy banks, in contrast, are unlikely to perceive any benefits from membership. If this happens, the funding for resolving problems will be questionable for both *ex-ante* and *ex-post* schemes. Indeed, the entire scheme may simply become a government bailout for weak banks. By requiring all banks to become members, the funding base is broader and more reliable. At the same time, when the healthy banks are members, they have a greater incentive to monitor and police the weaker banks to help protect the fund.

2.6 Lessons Learned from Banking Crises

It is quite clear that although many countries at all levels of income and in all parts of the world have established deposit insurance schemes, they have not chosen a uniform structure. The specific design features differ widely among the 133 countries for which complete information is available, as already discussed and indicated in Table 1. The fact that so many countries around the globe have suffered banking crises over recent years has generated a substantial amount of research focusing on the relationship between a banking crisis and deposit insurance. Although this type of research is still ongoing, there are currently enough studies from which to draw some, albeit tentative, conclusions about deposit insurance schemes that help promote a safe and sound banking industry. These are as follows:

- Even without a deposit insurance scheme, countries have on occasion responded to banking crises with unlimited guarantees to depositors. An appropriately designed scheme that includes a coverage limit may be better able to serve notice to depositors as to the extent of their protection, and thereby enable governments to avoid more costly *ex-post* bailouts.
- The design features of a deposit insurance scheme are quite important. Indeed, recent empirical studies show that poorly designed schemes increase the likelihood that a country will experience a banking crisis.
- Properly designed deposit insurance schemes can help mobilize savings in a country, and thereby help foster overall financial development. Research has documented this important linkage but emphasizes that it only holds in countries with a strong legal and regulatory environment.
- Empirical research shows that market discipline is seriously eroded in countries that have designed their deposit insurance schemes with a high coverage limit – an *ex-ante* fund – the government being the sole source of funds, and only public officials as the administrators of the fund.

- Empirical research shows that market discipline is significantly enhanced in countries that have designed their deposit insurance schemes with coinsurance, mandatory membership, and private or joint administration of the fund.

All in all, empirical research that has recently been completed indicates that governments should pay close attention to the features they wish to include in a deposit insurance scheme should they decide to adopt one or to modify the one they have already established (Barth et al. 2006).

2.7 Conclusions

Countries everywhere have shown a greater interest in establishing deposit insurance schemes in the past two decades. The evidence to date indicates that much more consideration must be given to the design features of these schemes to be sure that their benefits are not offset by their associated costs. This is important because, according to Anginer et al. 2019, “Deposit insurance coverage has increased and has also become more generous, especially for high-income OECD countries. This has likely encouraged banks – especially large banks – to take on excessive risk. Moreover, the information set available to assess the risk profile of banks did not improve significantly. . . .”

References

- Anginer, Deniz, Asli Demirguc-Kunt, and Min Zhu. 2014. How does deposit insurance affect bank risk? Evidence from the recent crisis. *Journal of Banking & Finance* 48: 312–321.
- Anginer, Deniz, Ata Can Bertay, Robert Cull, Asli Demirguc-Kunt, and Davide S. Mare. 2019. Bank regulation and supervision ten years after the global financial crisis. Policy Research Working Paper 9044, World Bank.
- Barth, James R., Gerard Caprio, and Ross Levine. 2006. *Rethinking bank regulation: Till angels govern*. Cambridge: Cambridge University Press.
- Bradley, Christine M. 2000. A historical perspective on deposit insurance coverage. *FDIC-Banking Review* 13 (2): 1–25.
- Cooper, Russell, and Thomas W. Ross. 2002. Bank runs: Deposit insurance and capital requirements. *International Economic Review* 43 (1): 55–72.
- Demirguc-Kunt, Asli, Bybars Karacaovali, and Luc Laeven. 2005. Deposit insurance around the world: A comprehensive database. The World Bank Policy Research Working Papers, 1(1).
- Demirguc-Kunt, Asli, and Enrica Detragiache. 2000. Does deposit insurance increase banking system stability? *International Monetary Fund Working Paper* WP/00/03.
- Demirguc-Kunt, Asli, and Edward Kane. 2002. Deposit insurance around the globe: Where does it work? *Journal of Economic Perspectives* 16 (2): 178–195.
- Demirguc-Kunt, Asli, and Tolga Sobaci. 2001. Deposit insurance around the world. *The World Bank Economic Review* 15 (3): 481–490.
- Demirguc-Kunt, Asli, Edward Kane, and Luc Laeven. 2015. Deposit insurance around the world: A comprehensive analysis and database. *Journal of Financial Stability* 20: 155–183.
- Federal Deposit Insurance Corporation. 2017. *Crisis and response: An FDIC history, 2008–2013*. Washington, DC: FDIC.

- Garcia, Gillian. 1999. Deposit insurance: A survey of actual and best practices. *International Monetary Fund Policy Working Paper* WP/99/54.
- . 2000. Deposit insurance and crisis management. *International Monetary Fund Policy Working Paper* WP/00/57.
- Gropp, R., and Jukka Vesala. 2004. Deposit insurance, moral hazard and market monitoring. *Review of Finance* 8 (4): 571–602.
- Gropp, Reint, Christian Gruendl, and Andre Guettler. 2014. The impact of public guarantees on bank risk-taking: Evidence from a natural experiment. *Review of Finance* 18 (2): 457–488.
- International Association of Deposit Insurers. 2020. Deposit Insurance Surveys 2020 Annual Survey. Accessed April 15, 2021. <https://www.iadi.org/en/core-principles-and-research/deposit-insurance-surveys/>
- Lee, Shih-Cheng, Chien-Ting Lin, and Ming-Shann Tsai. 2015. The pricing of deposit insurance in the presence of systematic risk. *Journal of Banking & Finance* 51: 1–11.
- Ngalawa, Harold, Fulbert Tchana Tchana, and Nicola Viegi. 2016. Banking instability and deposit insurance: The role of moral hazard. *Journal of Applied Economics* 19 (2): 323–350.
- Prescott, Edward S. 2002. Can risk-based deposit insurance premiums control moral hazard? *Federal Reserve Bank of Richmond Economic Quarterly* 88 (2): 87–100.
- The World Bank. 2019. Bank Regulation and Supervision Survey. Accessed April 15, 2021. <https://www.worldbank.org/en/research/brief/BRSS>
- . 2020. The global financial development report 2019/2020: Bank regulation and supervision a decade after the global financial crisis.



Gramm-Leach-Bliley Act: Creating a New Bank for a New Millennium

3

James R. Barth and Shen Zhang

Abstract

The Gramm-Leach-Bliley Act of 1999 repealed the Glass-Steagall Act of 1933 that had mandated the separation of commercial banking activities from securities activities. It also repealed provisions of the Bank Holding Company Act of 1956 that provided for the separation of commercial banking from insurance activities. The result was the establishment of a new legal structure that allowed for the integration of banking, securities, and insurance activities within a single organization. The key features of the Gramm-Leach-Bliley Act are explained and discussed, with special emphasis on its importance for US banks in a world of ever-increasing competition and globalization of financial services.

Keywords

Banking laws · Bank regulations · Financial holding companies · Financial modernization · Glass-Steagall · Globalization · Insurance · Securities · Thrifts

J. R. Barth (✉)

Auburn University and Milken Institute, Auburn, AL, USA

Department of Finance, Harbert College of Business, Auburn University, Auburn, AL, USA

e-mail: barthjr@auburn.edu

S. Zhang

Troy University, Troy, AL, USA

e-mail: zhangs@troy.edu

3.1 Introduction¹

The Gramm-Leach-Bliley Act (GLBA) was signed into law on November 12, 1999, and provided for sweeping changes in the allowable activities of banks in the USA (Barth et al. 2000). The GLBA, also known as the Financial Modernization Act, repealed the Glass-Steagall Act (GSA) of 1933 that had mandated the separation of commercial banking activities from securities activities. Also, the GLBA repealed provisions of the Bank Holding Company Act (BHCA) of 1956 that provided for the separation of commercial banking activities from insurance activities. While the GLBA formally changed the allowable financial activities of banks, the regulatory environment had been evolving since the 1980s, if not earlier, toward a less stringent interpretation of the GSA. In fact, in 1987 the Congressional Research Service, which provides policy and legal analysis to the US Congress, reviewed the arguments for and against preserving the GSA in response to the changing environment.²

The importance of the GLBA is that it removed years of restrictions on the mixing of banking, securities, and insurance activities within a single financial organization. When the GSA was enacted during the Great Depression following the stock market crash of 1929, the view of many was that speculation and conflicts of interest contributed to the crash and subsequent Great Depression (Benston 1990). The purpose of the GSA was to restore public confidence in the financial sector by delinking the banking system and securities markets. However, its restrictive provisions were gradually relaxed over the years and even more rapidly since the 1980s. The enactment of the GLBA is viewed by many as merely serving to formalize what had already been happening *de facto* in the financial marketplace as banks were allowed, albeit to a limited degree, to engage in a broader range of financial activities. A particularly defining event was when the Federal Reserve granted Citigroup in 1998 a temporary waiver of the GSA to allow the merger between Citicorp, a commercial bank holding company, and Travelers Group, an insurance company, to form the conglomerate Citigroup.

An additional motivation for the GLBA at the time was globalization. US banks had operated for decades under some of the most restrictive regulations, including those regarding the allowable range of financial activities, when compared to banks in most of the other industrial countries around the world. Even though the GLBA improved the position of banks in terms of global competitiveness, US banks still do not enjoy the same degree of freedom concerning activities and organizational structure as banks in many other countries.

¹ This chapter is based on an earlier version that was published and coauthored with John Jahera in 2013.

² See <https://web.archive.org/web/20100113065324/http://digital.library.unt.edu/govdocs/crs/permalink/meta-crs-9065:1>, accessed March 27, 2021.

3.2 Major Provisions of the Gramm-Leach-Bliley Act

3.2.1 Financial Holding Companies

The GSA and the BHCA severely restricted bank and/or bank subsidiary affiliations with securities firms and insurance companies. Figure 1 shows the allowable activities and the organizational structure under which banks operated before and after the new provisions provided by the GLBA.

The new law repealed earlier financial activity restrictions by creating new financial holding companies that are allowed through separately capitalized subsidiaries to engage in a wider range of activities if the Federal Reserve determines that the activities do not pose a substantial risk to bank safety or soundness.

The GLBA allows a bank holding company to become a financial holding company provided its subsidiary depository institutions are adequately capitalized, properly managed, and has a “satisfactory” rating under the Community Reinvestment Act (CRA). The new holding companies may engage in activities deemed to be financial in “nature” or “incidental” to financial activities. The Federal Reserve may also allow activities termed “complementary” to financial activities after determining that the activity does not impair the safety or soundness of banks. One caveat is that the Federal Reserve may not determine an activity to be financial in nature if the US Treasury Department objects. This provision may result in disputes regarding the interpretation of the law and hence add to uncertainty regarding approval of certain activities for banks. Over the years, however, this feature of the law has not been a source of problems.

Financial holding companies may own banks as subsidiaries as well as other subsidiaries that engage in other approved financial activities. Activities that the GLBA specifies to be “financial in nature” include underwriting and dealing in securities, insurance underwriting, and agency activities, merchant banking, mutual fund sponsorship, and insurance company portfolio investments. Insurance agency activities are regulated solely by the individual states and therefore may face state-imposed restrictions. However, states are precluded from restricting any activity that is specified in the GLBA.

3.2.2 National Bank Financial Subsidiaries

The GLBA also creates new financial subsidiaries of national banks (and, subject to state law, of state banks) that may engage in all the financial activities authorized by the new law. Exceptions include insurance or annuity underwriting, insurance company portfolio investments, real estate investment and development, and merchant banking. These latter activities may only be conducted in financial holding company subsidiaries. Furthermore, there is a limitation of the total assets of all financial subsidiaries to 45% of the total assets of the bank, or \$50 billion.

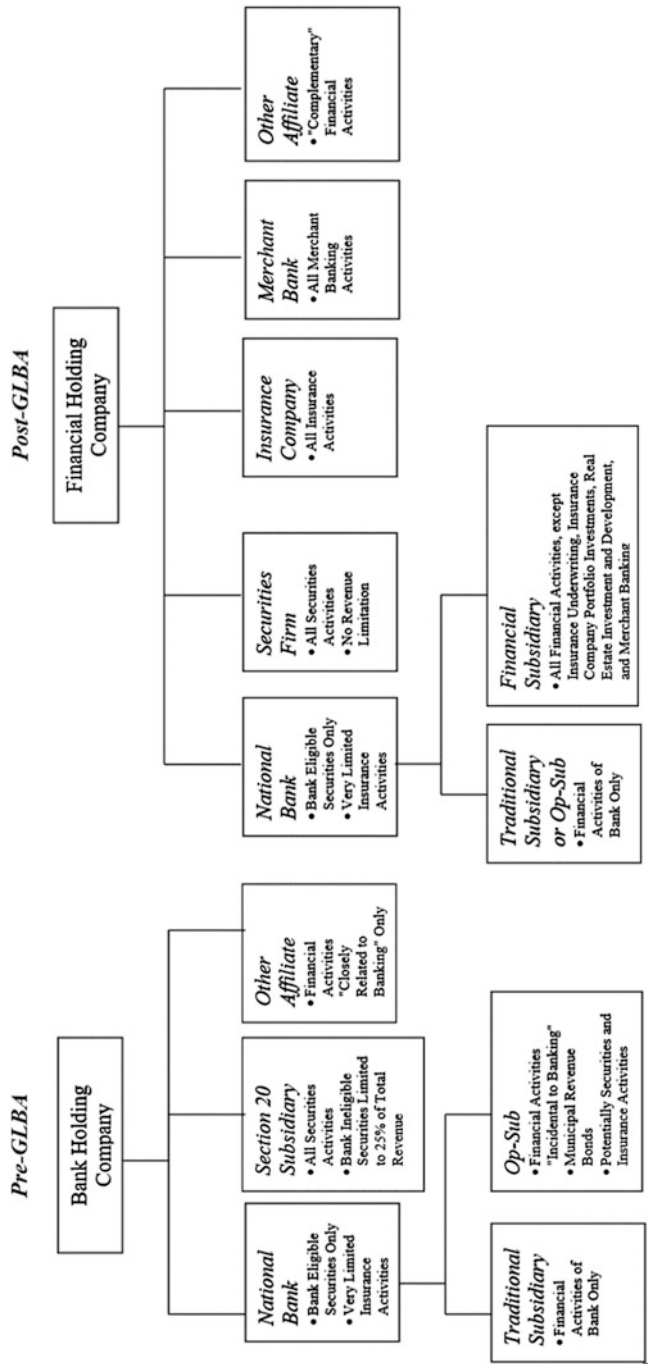


Fig. 1 Past and new bank permissible activities and organizational structure

3.3 Functional Regulation and Equal Treatment for Foreign Banks

The GLBA generally adheres to the principle of functional regulation, which holds that similar activities should be regulated by the same regulator. Thus, banking regulators regulate bank activities, securities regulators regulate securities activities, and insurance regulators regulate insurance activities. The Federal Reserve, as an umbrella regulator, is authorized to examine financial holding companies and their subsidiaries but may specifically examine functionally regulated subsidiaries only under limited circumstances. For those entities, the Federal Reserve will generally rely upon the examinations by other federal and state securities and insurance regulatory authorities.

The new law does exempt some banking activities that are deemed to have a “securities” component from the regulatory authority of the Securities and Exchange Commission (SEC). However, the law provides a process that requires the SEC to act by rulemaking before seeking to regulate any bank sales of any new hybrid security product. Finally, if a new product is determined to be an insurance product by a state insurance regulator, then national banks are prohibited from providing it as principal in that state. Any conflicts among the regulatory authorities must be resolved in a court of law, which has not been an issue to date.

The GLBA provides for national treatment whereby foreign banks may engage in the newly authorized financial activities on the same basis as domestic banking organizations.

3.3.1 Retention of Savings and Loan Holding Companies

The new law retained the federal savings and loan charter and allowed savings and loan holding companies (SLHCs) to conduct banking, securities, and insurance activities on the same terms as bank holding companies. The law, however, closed a loophole permitting the mixing of banking and commerce by prohibiting savings and loan holding companies from acquiring commercial firms or engaging in new commercial activity (Barth and Sun 2020). Over time, however, savings and loan institutions became like commercial banks in terms of the financial services offered. For this and other reasons, the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act) of 2010 transferred to the Federal Reserve the supervisory functions of the Office of Thrift Supervision (OTS) related to SLHCs and their nondepository subsidiaries as of July 21, 2011. Also, savings and loan institutions became subject to regulation by the bank regulatory authorities and members of the Federal Deposit Insurance Corporation (FDIC). This has led to the elimination of any meaningful distinctions between the two types of depository institutions.

The law also created new community financial institutions, which could obtain long-term Federal Home Loan Bank advances for lending to small businesses, small farms, and small agribusinesses. These institutions must be FDIC-insured depository institutions with less than \$500 million in assets. Thus, the US Congress provided

new government-directed subsidized lending to selected institutions to induce them to provide credit to businesses favored by it.

3.3.2 Community Reinvestment Act Provisions

The CRA of 1977 was enacted to ensure that banks do not lend the deposits gathered from individuals in one area to those living in another area in significant proportions. It thus requires banks to make credit available in the communities in which they obtain deposits. Before passage, there was concern that the CRA would be weakened. The GLBA, therefore, required that financial holding companies could not be formed until their insured depository institutions received and maintained a satisfactory CRA rating. Smaller institutions were granted some relief with less frequent CRA examinations. Banks with less than \$250 million in assets were to undergo a CRA examination once each 5 years if they have prior outstanding ratings and once every 4 years if they have prior satisfactory ratings. The GLBA further required that banks and community groups must disclose certain CRA agreements and provide annual reports on the use of funds and resources utilized in fulfilling such agreements. Financial holding companies and banks with financial subsidiaries are prohibited from new activities or acquisitions unless each insured institution within the company has earned at least a “satisfactory” CRA rating.

3.3.3 Other Components of GLBA

Given the growing importance of automated teller machines that charge fees, it was required that they be labeled with a notice of the fee. Machines must also give customers a notice on the screen that a fee will be charged, with the option of canceling the transaction.

The new law also required the relevant regulators to establish standards for ensuring the privacy of consumers’ personal information maintained by financial institutions. Surprisingly, whether consumer privacy would be adequately protected with the expansion of bank powers dominated congressional negotiations just before enactment. Indeed, in the House of Representatives, a shift in only 13 votes on the privacy provisions would have defeated the entire legislation.

Given this concern, the law required regulators to establish standards to ensure the privacy of personal financial information held by financial institutions. Also, consumers must be presented the opportunity to “opt out” of having their financial information shared with nonaffiliated third parties. Further, mandatory disclosure of the institution’s privacy policies must be made on an annual basis to all customers. One other feature designed to benefit consumers was the mandate that federal banking agencies must use “plain” language in all rules after January 1, 2000

[Banerji, Chen and Mazumdar (2002), Broome and Markham (2000), Carow and Heron (2002), and Wilmarth (2002)].

3.3.4 Potential Benefits to Banks and Their Customers

Banks potentially benefit from the expanded range of permissible activities through higher average profits resulting from scale and scope economies (Mester 2010; Boot 2016; Hughes et al. 2019). The fixed overhead cost of operating a bank may decrease as its size increases. Also, there may be economies when costs associated with overhead in administration, back-office operations, and information technologies are spread over a broader range of financial services that are offered.

Due to greater opportunities for diversification, a bank with broader powers may also have lower profit variability than a traditional bank. If profits from different financial activities are not highly, or even negatively, correlated, then the total profits of a broad bank will be more stable than that of banks specializing in relatively few activities. Broad banks, moreover, may be affected less when some firms bypass banks and raise funds directly in the capital markets because the decline in their lending activity may be offset by an increase in their securities activity. Also, customers may benefit from a broad bank. If a bank achieves greater scale and scope economies, for example, competition may lead to a sharing of these benefits with customers and firms in the form of lower prices. Furthermore, customers may benefit from lower search and transaction costs because of “one-stop” or “one-click” shopping.

Evaluating the welfare impact of the GLBA on customers is complex and may be subject to different assumptions and model specifications. To circumvent this obstacle, Filson and Olfati (2014) employ an indirect approach that estimates the GLBA diversification impact on the value of diversifying bank holding companies. They show that diversification into investment banking, securities brokerage, and insurance under the GLBA is associated with cumulative abnormal returns, suggesting such diversification creates value and, likely, enhances social welfare.

3.3.5 Potential Risk Elements to Banks and Their Customers

Two main concerns arise, however, when combining banking, securities, and insurance activities within the same banking organization. There may be a contagion effect insofar as problems in one unit may spill over and adversely affect other units. The greater range of activities, moreover, may increase the risk of insolvency to the organization. This might happen if banking organizations encounter unexpected difficulties in nontraditional activities, due to a lack of banks’ business experience or increased operational complexities or because the regulatory authorities might be less able to contain excessive risk-taking in the new activities.

Whether the expansion of securities and insurance powers puts banking organizations at greater risk of insolvency is an important issue. Bank regulators have

generally concluded that any risks associated with broader powers can be contained. According to Gary Whalen (1997), in a study conducted before the enactment of the GLBA, the “[o]rganizational structure and firewalls, in conjunction with supervision, appear capable of mitigating concerns about increases in risk stemming from broader bank powers.” Also, the FDIC supported the repeal of the GSA on the grounds it would advance financial modernization without sacrificing safety and soundness (Barth et al. 2004). More recently, the severe problems in the banking industry during the housing boom and bust during 2000–2007 and the subsequent Great Recession during 2007–2009 were not attributed to the broader powers granted by the GLBA.

The federal safety net problem, either explicit or implicit, is the second concern. The explicit safety net refers to extending the benefits of federal deposit insurance and access to both the payment system and the discount window of the Federal Reserve to a broader range of activities. The implicit part refers to the broadly held perception that some systemically important financial institutions are “too big to fail” (Kaufman (2014), Barth and Wihlborg (2017), and Cetorelli and Traina (2018)). That is, the government would provide financial support to such institutions when they are at or near insolvency to mitigate substantial losses borne by creditors and investors. The implicit part becomes increasingly more important as the banking industry has consolidated and become more concentrated post-GLBA (see Table 1). If banks receive a subsidy from access to the federal safety net and if it can be extended to additional activities, then banks may possess an unfair advantage vis-à-vis their nonbank competitors in these activities. Furthermore, such a situation might encourage banks to engage too heavily in additional activities and thereby become riskier.

Banks, however, also incur special costs associated with the federal safety net. They pay premiums for deposit insurance, have historically held interest-free

Table 1 US commercial banking structure, 1979–2019

Year	Assets \$ Trillion	Number of banks	Number of branches	Number of offices	Banks with branches	Unit banks	New charters
1979	1.7	14,364	36,791	51,155	6,564	7,800	204
1984	2.5	14,496	41,625	56,121	7,094	7,402	356
1989	3.3	12,715	47,789	60,504	6,972	5,743	170
1994	4.0	10,453	54,384	64,837	6,649	3,804	46
1999	5.7	8,582	63,631	72,213	5,972	2,610	228
2004	8.4	7,628	70,313	77,941	5,590	2,038	120
2009	11.8	6,829	82,490	89,319	5,225	1,604	24
2014	14.5	5,607	81,405	87,012	4,446	1,161	0
2019	17.5	4,518	76,837	81,355	3,680	838	13

Note: This table shows the banking structure in the USA from 1979–2019. Number of offices is the number of banks plus the number of branches. Unit Banks is the number of banks minus banks with branches. Source: [BankFind \(fdic.gov\)](https://www.fdic.gov/bankfind/), accessed April 2021

reserves,³ and incur costs to satisfy numerous banking rules and regulations. These costs must be subtracted from any gross subsidy that some banks may receive to obtain the net subsidy. In this regard, Dahl, Meyer, and Neely (2016) show banks' regulatory compliance costs range from \$164 thousand dollars to \$1.8 million dollars, depending on the size of the bank. This represents 8.7% of the overall noninterest expenses for banks with assets less than \$100 million and 2.9% for banks with assets between \$1 billion and \$10 billion. Despite these costs, Luc Laeven (2016) calculates net subsidies for banks were 0.4 basis points in the USA.

3.3.6 Implications for the Future

The World Bank has conducted five surveys on bank regulations and supervisions since 2000 and received responses from more than 100 countries (World Bank Group 2020; Anginer et al. 2019). According to the most recent survey in 2019, information is available for all 32 nonoverlapping G-10 and EU countries (including the UK) on the restrictions, if any, placed on banks engaging in securities, insurance, and real estate activities. As Table 2 shows, 29 of the countries allow a full range of securities activities to be conducted directly in the bank, while the remaining three require that the activities be conducted in subsidiaries or another part of a common holding company or parent. As regards insurance activities, 28 countries allow a full range of activities but only 2 allow them to be directly conducted in the bank, with the remaining 26 requiring the activities to be conducted in subsidiaries or another part of a common holding company or parent. Three of the countries prohibit a bank from engaging in insurance activities, while one only allows less than a full range of activities. The table also provides information on the extent to which the same countries allow banks to engage in real estate activities. These activities are somewhat more restricted than either securities or insurance activities. When combining the restrictions placed on all three activities, the USA ranks in the top six in terms of overall restrictiveness.

Importantly, in an analysis of more than 107 countries, Barth, Caprio, and Levine (2004) found that tighter restrictions placed on bank activities had negative repercussions. In particular, they found that greater regulatory restrictions on bank activities in a country are associated with (1) a higher probability of that country suffering a major banking crisis and (2) lower banking sector efficiency in the country.

By permitting banks to engage in banking, securities, and insurance activities, and by providing even broader powers to financial holding companies, the GLBA helped rejuvenate banking since its passage. While banks held nearly three-fourths of the total assets of all financial intermediaries in 1860, recently their share had declined to less than one-fourth. The combined assets of commercial banks,

³On October 6, 2008, the Federal Reserve Board announced that it would begin paying interest on depository institutions' both required reserve balances and excess reserve balances.

Table 2 Regulatory restrictions on bank activities in selected countries

	Securities	Insurance	Real estate	Total score
Austria	1	2	1	4
Belgium	1	2	1	4
Bulgaria	1	2	2	5
Canada	2	2	2	6
Croatia	1	3	2	6
Cyprus	1	2	2	5
Czech	1	2	2	5
Denmark	1	2	2	5
Estonia	1	2	1	4
Finland	1	2	2	5
France	1	2	1	4
Germany	1	2	1	4
Greece	1	2	2	5
Hungary	1	2	2	5
Ireland	1	2	4	7
Italy	1	2	3	6
Japan	2	2	4	8
Latvia	1	1	2	4
Lithuania	1	4	4	9
Luxembourg	1	2	2	5
Malta	1	3	4	8
Netherlands	1	2	1	4
Poland	1	2	2	5
Portugal	1	2	2	5
Romania	1	2	1	4
Slovakia	1	4	4	9
Slovenia	1	2	2	5
Spain	1	2	2	5
Sweden	1	2	2	5
Switzerland	1	1	1	3
UK	1	4	1	6
USA	2	2	3	7

Note: The table indicates the extent to which restrictions are placed on banks engaging in securities activities, insurance activities, real estate activities, and nonfinancial businesses. 1 indicates that a full range of these activities can be conducted directly in banks; 2 indicates that a full range of these activities is allowed but all or some of these activities must be conducted in subsidiaries, or in another part of a common holding company or parent; 3 indicates that less than the full range of activities can be conducted in banks, or subsidiaries, or in another part of a common holding company or parent; 4 indicates that none of these activities can be done in either banks or subsidiaries, or in another part of a common holding company or parent

insurance companies, securities firms, and investment companies have decreased as other financial firms have taken advantage of new technology and grown in importance.

Capital markets (stocks and bonds) are now far more important in countries as compared to bank loans than in the last century. This shift in the composition of the financial system reflects the fact that financial intermediation based upon a securities-based system has become more cost-effective and thus more competitive with a bank loan-based system. As of December 31, 2020, the cost of intermediation through a bank was about 268 basis points as measured by net interest margin. This compares to less than 100 basis points as measured by the operating expense ratio of mutual funds.

These newer developments have forced banks to rely less on traditional net interest income and more on fee-based income. Indeed, noninterest income as a percentage of net operating revenue, as of December 31, 2020, was 40% for banks with more than \$250 billion in assets and around 30% for banks with less than \$250 billion in assets. The emphasis of banks is increasingly on asset and risk management, especially for the biggest banks. The traditional definition of a bank, moreover, was a firm that made commercial and industrial loans, accepted demand deposits, and offered federally insured deposits. Although banks are still insured by the FDIC, commercial and industrial loans were only 11% of total assets, while demand deposits were 21%, at year-end 2020.

3.4 Conclusion

Traditional banks, therefore, are now largely the banks of the past. The newer broad banks reflect the historic changes brought about mainly by technology and globalization, as well as the corresponding regulations facilitating these developments. Providing liquidity in the form of deposits and loans to businesses will undoubtedly remain an important service of some community banks, but it will be subsumed in the broader strategy of asset and risk management for bigger banks. Modern information technology, moreover, will enable all banks to rely less on a branch network to attract and maintain customers. At the same time, however, banks of all sizes will face increasing competition on both sides of their balance sheets from fintech firms and digital currencies.

References

- Anginer, Deniz, Ata Can Bertay, Robert J. Cull, Asli Demirguc-Kunt, and Davide S. Mare. 2019. *Bank regulation and supervision ten years after the global financial crisis*. Policy research working Paper 9044. World Bank, Washington, DC.
- Banerji, Sanjay, Andrew H. Chen, and Sumon C. Mazumda. 2002. Universal banking under bilateral information asymmetry. *Journal of Financial Services Research* 22 (3): 169–187.
- Barth, James R., and Yanfei Sun. 2020. Industrial banks: Challenging the traditional separation of commerce and banking. *The Quarterly Review of Economics and Finance* 77: 220–249.
- Barth, James R., and Clas Wihlborg. 2017. Too big to fail: Measures, remedies, and consequences for efficiency and stability. *Financial Markets, Institutions & Instruments* 26 (4): 175–245.
- Barth, James R., R. Dan Brumbaugh Jr., and James A. Wilcox. 2000. Policy watch: The repeal of Glass-Steagall and the advent of broad banking. *Journal of Economic Perspective* 14 (2): 191–204.

- Barth, James R., Gerrard Caprio, and Ross Levine. 2004. Bank regulation and supervision: What works best? *Journal of Financial Intermediation* 13 (2): 205–248.
- Benston, George J. 1990. *The separation of commercial and investment banking: The Glass-Steagall act revisited and reconsidered*. Basingstoke: Oxford University Press.
- Boot, Arnoud W.A. 2016. Understanding the future of banking: Scale & scope economies, and Fintech. In *The future of large internationally active banks*, ed. Asli Demirgüç-Kunt, Douglas D. Evanoff, and George G. Kaufmann, vol. 55, 429–448. New Jersey: World Scientific Studies in International Economics.
- Broome, Lissa L., and Jerry W. Markham. 2000. Banking and insurance: Before and after the Gramm-Leach-Bliley Act. *Journal of Corporation Law* 25 (4): 723–785.
- Carow, Kenneth A., and Randall A. Heron. 2002. Capital market reactions to the passage of the Financial Services Modernization Act of 1999. *The Quarterly Review of Economics and Finance* 42 (3): 465–485.
- Cetorelli, Nicola, and James Traina. 2018. “Resolving “Too Big to Fail.” Federal Reserve Bank of New York, Staff Report No. 859.
- Dahl, Drew, Andrew Meyer, and Michelle Neely. Scale matters: Community banks and compliance costs. Federal Reserve Bank of St. Louis, *Regional Economist*, July 2016.
- Filson, Darren, and Saman Olfati. 2014. The impacts of gramm–leach–bliley bank diversification on value and risk. *Journal of Banking and Finance* 41: 209–221.
- Hughes, Joseph P., Julapa Jagtiani, Loretta J. Mester, and Choon-Geol Moon. 2019. Does scale matter in community bank performance? Evidence obtained by applying several new measures of performance. *Journal of Banking and Finance* 106: 471–499.
- Kaufman, George G. 2014. Too big to fail in banking: What does it mean? *Journal of Financial Stability* (13): 214–223.
- Laeven, Luc A. 2016. *Banking risks around the world: The implicit safety net subsidy approach*. Working Paper. Available at SSRN: <https://ssrn.com/abstract=632543>
- Mester, Loretta J. 2010. Scale economies in banking and financial regulatory reform. Federal Reserve Bank of Minneapolis.
- Whalen, Gary. 1997. *Bank organizational form and the risks of expanded activities*. Office of the Comptroller of the Currency, Administrator of National Banks 97 (1).
- Wilmarth, Arthur E., Jr. 2002. The transformation of the U.S. financial services industry, 1975–2002: Competition, consolidation and increased risks. *University of Illinois Law Review* (2): 215–476.
- World Bank Group. 2020. *Bank regulation and supervision a decade after the global financial crisis*, Global financial development report. Washington, DC: World Bank Group.

Pre-funded Coupon and Zero-Coupon Bonds: Cost of Capital Analysis

4

Suresh Srivastava and Ken Hung

Abstract

Pre-funded coupon bonds have been developed and sold by investment bankers in place of zero-coupon bonds to raise funds for companies facing cash flow problems. Additional bonds are issued and proceeds are deposited in an escrow account to finance the coupon payment. Our analysis indicates that a pre-funded coupon bond is equivalent to a zero-coupon bond only if the return from the escrow account is the same as the yield to maturity of the pre-funded issue. In reality, the escrow return is lower than the bond yield. As a result, the firm provides interest subsidy through issuing additional bonds which leads to higher leverage, greater risk and loss of value compared to a zero-coupon issue.

Keywords

Financial engineering · Macaulay duration · Zero-coupon bond

4.1 Introduction

Pre-funded coupon bonds were first issued in 1994 (Doherty 1997). They were introduced as a means to raise capital for firms unable to generate cash-flow to make coupon payments, while still meeting the needs of investors to receive coupon income. With a pre-funded bond structure, additional bonds are issued and an escrow account is established to finance coupon payments over the life of the bond. In this manner, the bond is considered pre-funded. The firm is not required to generate cash

S. Srivastava (✉)

University of Alaska Anchorage, Anchorage, AK, USA

e-mail: afscs@uaa.alaska.edu

K. Hung

Division of International Banking and Finance Studies, Texas A&M International University,
Laredo, TX, USA

flow to meet coupon obligations; it is paid out of the escrow account usually collateralized by zero-coupon Treasury securities. The risk free coupon payment allows the firm to set a lower coupon rate on the bond than the yield on a comparable zero-coupon bond. In general, the cost of funding the escrow account is greater than the return of the escrow account. This leads to an interest rate subsidy and the loss of value. In this chapter, we compare zero-coupon bonds to pre-funded bonds and ascertain conditions under which the two the funding options are equivalent. A pre-funded issue simultaneously creates an asset and a liability. The net duration of the pre-funded issue is the weighted average of the asset and liability durations. The model of net duration developed in this chapter incorporates increased leverage of the pre-funded issue and appropriately assess its increased risk. In spite of the fact that pre-funded bond is an interesting concept of financial engineering, there is very little academic research on this topic.

The remainder of this chapter is made up of four sections. Section 4.2 discusses the options available to a firm interested in issuing debt. In Sect. 4.3 we derive a mathematical model for McCauley duration of the pre-funded issue to determine the interest rate risk and calculate the loss in value due to interest rate subsidy. A numerical example and its analysis are presented in Sects. 4.4 and 4.5 concludes the paper.

4.2 Funding Options

A firm wants to raise funds to finance a new project. The pecking-order theory of capital structure suggests that managers prefer internal equity to external financing (Myers 1984). In case the internal equity (retained earnings) is not available then issuing new debt is preferred over issuing preferred or additional common stock. Further, firm would like to reduce the interest payment burden. Hence, conventional coupon bond or hybrid financing such as convertible bonds or bonds with warrants are ruled out. The available funding options are (1) zero-coupon bonds, (2) step-up bonds – initially coupon payment is set at a low value and later stepped up, (3) deferred interest bonds – initially there is no interest payment, but it is resumed in 3–7 years, (4) paid-in-kind bonds – issuer has right to pay interest in cash or with similar bonds,¹ and (5) -pre-funded bonds. The focus of the study is to compare zero-coupon and pre-funded bonds.

4.2.1 Zero-Coupon Bonds

Pure discount bonds are often called zero-coupon bonds. It was first issued by J.C. Penney Company Inc. in 1982 (Brigham and Daves 2010). In recent years, other firms (e.g., IBM, GMAC, Alcoa and Martin-Marietta) have issued zero-coupon

¹See Goodman and Cohen (1989) for detailed discussion of paid-in-kind bonds.

bonds. Municipalities started issuing zero-coupon bonds in 1983. These bonds are sold at a deep discount and increase in value as they approach maturity. Zero-coupon bonds do not provide interest or coupon payments at regular intervals like other types of bonds. Implicit coupons are automatically reinvested by the issuer at yield to maturity. Interest accrues over the life of the bond and a return is earned as the bond appreciates. At maturity its value equals the face value and bond holder receives the yield to maturity expected at the time of purchase. If held to maturity, the investor faces no reinvestment risk but high interest rate risk, as its market price fluctuates considerably with movements in market rates.

Corporate and municipal zero-coupon are usually callable and rated as junk bonds. The financial condition of the company issuing bonds predicates the use of junk bonds, i.e., the firm is unable generate cash flows to meet coupon payments. Junk bonds are typically rated BB or lower by Standard and Poor's, or BA and lower by Moody's. Junk bonds offer a high expected return but require investors to take on higher default risk. Covenants on junk bonds are less restrictive and therefore provide alternatives for firms that may not meet the more restrictive covenants of conventional bonds. U.S. Treasury sells Zero-coupon bonds in the form of STRIPs (Separate Trading of Registered Interest and Principle of Security Program). A long term T-bond is stripped into its coupon payments and the principle. Each cash flow is sold as a zero-coupon bond. Lately, these zero-coupon Treasuries have become very popular with the investors because of low interest rate and inflation (Zeng 2008).

4.2.2 Pre-funded Bonds

In raising capital with a pre-funded bond issue, additional bonds are issued and an escrow account is established.² The firm is not required to generate cash flow to meet coupon obligations over the life of the bond. Bond interests are paid out of an escrow account, which is usually collateralized by Treasury securities. In this manner, the bond is considered pre-funded. A pre-funded bond issue simultaneously creates an asset and a liability. The risk characteristics of pre-funded bond's interest payments are different from that of traditional coupon-bearing bonds because pre-funded bond's coupon payments are asset based. The default free nature of the coupon payment allows the firm to set a lower coupon rate than the yield on a comparable zero-coupon bond. In general, the cost of funding the escrow account is greater than the return from the escrow account. This spread leads to an interest rate subsidy which necessitates issuing more bonds and hence a loss of value. Greater the spread between cost of funding the escrow account and the return from the escrow account, larger the total face value of the pre-funded issue and the value loss. With a pre-funded bond issue there are additional flotation costs and cost of establishing

²Bierman (2005) cites some of the recent examples of bond interest collateralization and investor's perspective.

the escrow account. However, for this analysis, we consider the escrow costs and additional flotation costs to be negligible.

Market price of pre-funded bonds fluctuates with movements in market rates, but it does not move as dramatically as zero-coupon bond prices. The reason for this difference is that zero-coupon bonds do not provide any cash flow until maturity. Coupon payments reduce the impact of interest rate changes on pre-funded bonds. Market conditions where interest rate movements are frequent and highly variable make pre-funded bond more attractive than zero-coupon bonds. The risk profiles of zero-coupon and pre-funded bonds can be summarized as follows: A zero-coupon bond has no reinvestment risk, higher price elasticity to interest rate changes and a default risk consistent with its junk bond rating. The pre-funded bond has reinvestment risk but lower price elasticity to interest rate changes. For a meaningful analysis of the interest rate risk, one must examine the combined interest rate sensitivity of the escrow asset and the bond liability. The default risk of the pre-funded issue should be decomposed into two components: the default risk of the coupon payments and the default risk of the maturity payment. The coupon payments are default free but the default risk of the maturity payments are much higher. This is due to increased leverage of the pre-funded issue compared to zero-coupon financing. In spite of the default free coupon payments, the pre-funded bonds are usually rated as junk bonds.

In the next section, the combined interest rate sensitivity of the escrow asset and the bond liability is examined. A model for the net Macaulay duration of the pre-funded issue is developed and loss of value due to interest rate subsidy is calculated.

4.3 Mccauley Duration and Value Loss

In this section we calculate the total face value of the pre-funded bonds issued, initial balance of the escrow account, interest rate subsidy provided by the firm, effective cost of the pre-funded issue and resulting loss of value. Also, we derive an expression for the net Macaulay duration of the pre-funded issue, i.e., the weighted average durations of the coupon bond and the escrow asset.

The face values of zero-coupon bonds issued, to raise an amount B , is

$$B_z = B(1 + r_z)^n \quad (1)$$

where r_z is the discount rate for the zero-coupon bond with maturity n . The Macaulay duration of zero-coupon bond is its maturity (Fabozzi 2010).

Let B_{pf} be the face value of the pre-funded bonds issued to raise an amount B . The annual coupon payment is $B_{pf}(r_{pf})$, where r_{pf} is the pre-funded bond yield. The initial balance in the escrow annuity account set up to meet the coupon payments is $B_{pf} - B$. Hence,

$$B_{pf} - B = (B_{pf}r_{pf})(PVIFA_{r_{es},n})$$

Or

$$B_{pf} = \frac{B}{1 - r_{pf}(PVIFA_{r_{es},n})} \quad (2)$$

where PVIFA indicates present value interest factor of an annuity, n is the maturity and r_{es} is the rate of return on the escrow account. Substituting the algebraic expression for PVIFA we get³

$$B_{pf} = \frac{r_{es}(1 + r_{es})^n B}{r_{pf} - (r_{pf} - r_{es})(1 + r_{es})^n} \quad (3)$$

The initial balance in the escrow account is

$$B_{pf} - B = \frac{r_{pf}\{(1 + r_{es})^n - 1\}B}{r_{pf} - (r_{pf} - r_{es})(1 + r_{es})^n} \quad (4)$$

Escrow account is funded at a cost of r_{pf} and provides a return of r_{es} . Consequently, the firm is providing a pre-tax interest subsidy of $(r_{pf} B_{pf})(r_{pf} - r_{es})$ per year, which increases the cost of pre-funded issue and leads to loss of value.

The loss of value is:

$$\text{Value Loss} = (r_{pf} B_{pf})(r_{pf} - r_{es}) \frac{(1 + r_{pf})^n - 1}{(1 + r_{pf})^n} \quad (5)$$

and the effective cost of the pre-funded issue is given by:

$$r_{eff} = \left(\frac{r_{es}(1 + r_{es})^n}{r_{pf} - (r_{pf} - r_{es})(1 + r_{es})^n} \right)^{\frac{1}{n}} - 1 \quad (6)$$

The concept of duration was introduced by Macaulay (1938) as a measure of price sensitivity of an asset or liability to a change in interest rates. Working independently, Samuelson (1945) and Redington (1952) developed the same concept about the interest rate risk of bonds. Details of duration computation can be found in any finance text (Fabozzi 2010). A pre-funded bond issue creates an asset, the escrow account annuity with market value $B_{pf} - B$; and a liability, coupon bonds with market value B_{pf} . The net market value of the pre-funded issue is B . Let D_{es} and D_{pf} represent the duration of escrow annuity and the bond liability respectively. Duration D_{es} is the Macaulay duration of an n -year annuity with yield r_{es} and D_{pf} is the Macaulay duration of an n -year coupon bond with yield to maturity r_{pf} . The net duration of the pre-funded issue is the weighted average of the durations of the escrow account and the coupon bond. Hence

³See Ross et al. (2010) for algebraic expression of PVIFA.

$$D_{net} = \frac{B_{pf}}{B} x D_{pf} - \frac{B_{pf} - B}{B} x D_{es} \quad (7)$$

where $\{B_{pf}/B\}$ and $\{-(B_{pf} - B)/B\}$ are the weights of the coupon bond and the escrow annuity respectively. This definition of net duration, D_{net} , captures the increased risk due to additional leverage caused by pre funding of coupon payments and interest subsidy provided by the firm.

4.4 Numerical Example and Analysis

A firm wants to raise \$10 million by issuing either zero-coupon bonds or pre-funded bonds with 5 or 10 year maturity. We assume that transaction costs are identical for both issues and negligible (Alternately, we can assume that all yields are net of transaction costs). Further, we assume that financial market views the zero-coupon and pre-funded bonds to be equivalent securities, and prices them with identical yields. Four different yields, 8%, 7%, 6% and 5% for zero-coupon and pre-funded bonds are considered for this analysis. Later, we modify this assumption and consider the situation where market views pre-funded bond to be safer and erroneously prices them with yields lower than the comparable zero-coupon yields by 25, 50 and 75 basis points. In doing so, market overlooks the added default risk associated with increased leverage.

Table 1 presents the face value of zero-coupon bonds issued to meet \$10 million funding need. For 5-year maturity with discount rates of 8%, 7%, 6%, and 5%, the firm issues zero-coupon bonds with total face values of \$14,693,281, \$14,025,517,

Table 1 Zero-coupon bond

		Discount rate, r_z			
		8%	7%	6%	5%
Maturity, n 5 years	Funds needed, B	\$10,000,000	\$10,000,000	\$10,000,000	\$10,000,000
	Face value of bonds issued, B_z	\$14,693,281	\$14,025,517	\$13,382,256	\$12,762,816
	Market value of bonds issued	\$10,000,000	\$10,000,000	\$10,000,000	\$10,000,000
	Duration, D_z	5 years	5 years	5 years	5 years
Maturity, n 10 years	Funds needed, B	\$10,000,000	\$10,000,000	\$10,000,000	\$10,000,000
	Face value of bonds issued, B_z	\$21,589,250	\$19,671,514	\$17,908,477	\$16,288,946
	Market value of bonds issued	\$10,000,000	\$10,000,000	\$10,000,000	\$10,000,000
	Duration, D_z	10 years	10 years	10 years	10 years

$$B_z = B(1 + r_z)^n \text{ and } D_z = n$$

\$13,382,256 and \$12,762,816 respectively. These values are calculated using Eq. 1. The Macaulay duration of the 5-year zero-coupon bond is 5 years. For 10-year zero-coupon bonds, an 8%, 7%, 6%, and 5% discount rate leads to total face values of \$21,589,250, \$19,671,514, \$17,908,477 and \$16,288,946 respectively. The Macaulay duration of the 10-year zero-coupon bond is 10 years.

In Table 2 we present the total face value of the pre-funded issue, amount of annual coupon payment disbursed from escrow account and the effective cost of pre-funded issue. It provides following important inferences.

First, when the pre-funded bond yield, r_{pf} , is the same as the escrow account return, r_{es} , then (1) total face value of the pre-funded issued is the same as the total face value of the zero-coupon bonds and (2) the effective cost of pre-funded issue, r_{eff} , is the same as the yield to maturity of the zero-coupon bond, r_z . Second, increase in the spread between r_{pf} and r_{es} increases the total face value of the bonds issued and its effective cost. Finally, for a given spread the total face value of the bonds issued and its effective cost increases with maturity. For example, consider the case when both r_{pf} and r_{es} are equal to 8% and firm wants to issue bonds 5-year maturity bonds to raise \$10 million. It can issue either zero-coupon bonds or pre-funded coupon bonds with \$14,693,281 face value and 8% effective costs. For 10-year maturity, it will have to issue \$21,589,250 zero-coupon or pre-funded bonds. However, with a 3% spread, i.e., $r_{pf} = 8\%$ and $r_{es} = 5\%$, the firm will have to issue \$15,298,893 coupon bonds with maturity 5 years or \$26,160,132 coupon bonds with maturity 10 years. The effective cost of 5-year and 10-year pre-funded issues will rise to 8.876% and 10.094% respectively.

Examples of net duration of pre-funded issue, i.e., the weighted average durations of the escrow asset and coupon bond liability are presented in Tables 3 and 4. In Table 3 we present a 5-year bond issue without spread i.e., both r_{pf} and r_{es} are equal to 8%. Firm issues \$14,693,281 bonds with annual coupon payment of \$1,175,462. Coupon payments are disbursed out of an escrow account with \$4,693,281 initial balance. Panel A of Table 3 shows that duration of the coupon bond, D_{pf} , is 4.3121 years. Panel B of Table 3 shows that the duration of the escrow annuity, D_{es} , is 2.8465 years. Panel C of Table 3 shows that the weights of bond liability and escrow asset are 1.469 and -0.469 respectively. Hence, the net duration, D_{net} , of the pre-funded issue is 5 years, which is identical to the duration of a zero-coupon bond. The result is understandable because firm has no net cash outflow for years 1–4, the only cash outflow of \$14,693,281 is in year 5.

In Table 4 we present an example of 5-year pre-funded bond issue with 3% spread i.e., $r_{pf} = 8\%$ and $r_{es} = 5\%$. Firm issues \$15,298,250 bonds with annual coupon payment of \$1,223,912. Coupon payments are disbursed out of an escrow account with \$5,298,250 initial balance. Firm provides the interest subsidy by issuing additional bond compared to Table 3 example. Panel A of Table 4 shows that the duration of the coupon bond, D_{pf} , is 4.3121 years, same as the Table 3 example. But the duration of the escrow annuity, D_{es} , increases to 2.9025 years. The weights of bond liability and escrow asset, reported in Panel C of Table 4, are 1.530 and -0.530

Table 2 Total face value and effective cost of pre-funded issue $B_{pf} = \frac{r_{es}(1+r_{es})^n B}{r_{pf} - (r_{pf} - r_{es})(1+r_{es})^n}$ and

$$r_{eff} = \left(\frac{r_{es}(1+r_{es})^n}{r_{pf} - (r_{pf} - r_{es})(1+r_{es})^n} \right)^{\frac{1}{n}} - 1$$

	Escrow return, r_{es} (%)	Pre-funded bond yield, r_{pf}				
			8%	7%	6%	5%
Maturity, n 5 years	8	Face value, B_{pf}	\$14,693,281			
		Escrow payment	\$1,175,462			
		Effective cost, r_{eff}	8.000%			
	7	Face value, B_{pf}	\$14,881,302	\$14, 025,517		
		Escrow payment	\$1,190,504	\$981,786		
		Effective cost, r_{eff}	8.275%	7.000%		
	6	Face value, B_{pf}	\$15,082,708	\$14,181,691	\$13,382,256	
		Escrow payment	\$1,206,617	\$992,718	\$802,935	
		Effective cost, r_{eff}	8.567%	7.237%	6.000%	
	5	Face value, B_{pf}	\$15,298,893	\$14,368,507	\$13,509,289	\$12,762,816
		Escrow payment	\$1,223,912	\$1,004,395	\$810,557	\$638,141
		Effective cost, r_{eff}	8.876%	7.518%	6.201%	5.000%
Maturity, n 10 years	8	Face value, B_{pf}	\$21,589,250			
		Escrow payment	\$1,727,140			
		Effective cost, r_{eff}	8.000%			
	7	Face value, B_{pf}	\$22,825,137	\$19,671,514		
		Escrow payment	\$1,826,011	\$1,377,006		
		Effective cost, r_{eff}	8.603%	7.000%		

(continued)

Table 2 (continued)

	Escrow return, r_{es} (%)	Pre-funded bond yield, r_{pf}				
			8%	7%	6%	5%
	6	Face value, B_{pf}	\$24,319,478	\$20,627,322	\$17,098,477	
		Escrow payment	\$1,945,558	\$1,443,913	\$1,074,509	
		Effective cost, r_{eff}	9.294%	7.509%	6.000%	
	5	Face value, B_{pf}	\$26,160,123	\$21,763,801	\$18,632,525	\$16,288,946
		Escrow payment	\$2,092,810	\$1,523,466	\$1,117,952	\$814,447
		Effective cost, r_{eff}	10.094%	8.087%	6.421%	5.000%

Empty cell represent the improbable case of $r_{pf} < r_z$. If $r_{es} = r_{pf}$ then $B_{pf} = B(1 + r_{pf})^n = B_z$ and $r_{eff} = r_{pf}$

Table 3 Net duration of the pre-funded issue $D_{net} = \frac{B_{pf}}{B}xD_{pf} - \frac{B_{pf}-B}{B}xD_{es}$

Panel A: Bonds issued					
Time, t	Cash outflow, CF	PVIF _{8%, 5}	CF*PVIF	t*CF*PVIF	Duration, D_{pf}
1	\$1,175,462	0.9259	\$1,088,391	\$1,088,391	
2	1,175,462	0.8573	1,007,769	2,015,538	
3	1,175,462	0.7938	933,120	2,799,359	
4	1,175,462	0.7350	864,000	3,455,999	
5	15,868,743	0.6806	10,800,000	53,999,999	
			\$14,693,280	\$63,359,286	4.3121
Panel B: Escrow annuity					
Time, t	Cash inflow, CF	PVIF _{8%, 5}	CF*PVIF	t*CF*PVIF	Duration, D_{es}
1	\$1,175,462	0.9259	\$1,088,391	\$1,088,391	
2	1,175,462	0.8573	1,007,769	2,015,538	
3	1,175,462	0.7938	933,120	2,799,359	
4	1,175,462	0.7350	864,000	3,455,999	
5	1,175,462	0.6806	800,000	3,999,998	
			\$4,693,280	\$13,359,282	2.8465
Panel C: Net durations					
Fund raised, B	\$10,000,000	Escrow amount, $B_{pf} - B$		\$4,693,281	
Face value of bond, B_{pf}	\$14,693,281	Escrow return, r_{es}		8%	
Bond yield, r_{pf}	8.00%	Escrow weight, $(B_{pf} - B)/B$		0.469	
Bond weight, B_{pf}/B	1.469	Escrow duration, D_{es}		2.847	
Bond duration, D_{pf}	4.312	Net duration, D_{net}		5.000	

If escrow return equals the bond yield, i.e., $r_{es} = r_{pf}$, then the net duration equals the maturity

Table 4 Net duration of the pre-funded issue $D_{net} = \frac{B_{pf}}{B} \times D_{pf} - \frac{B_{pf} - B}{B} \times D_{es}$

Panel A: Bonds issued					
Time, t	Cash Outflow, CF	PVIF _{8%, 5}	CF*PVIF	t*CF*PVIF	Duration, D _{pf}
1	\$1,223,912	0.9259	\$1,133,252	\$1,133,252	
2	1,223,912	0.8573	1,049,307	2,098,615	
3	1,223,912	0.7938	971,581	2,914,742	
4	1,223,912	0.7350	899,612	3,598,447	
5	16,522,162	0.6806	11,244,706	56,223,529	
			\$15,298,458	\$65,968,585	4.3121
Panel B: Escrow annuity					
Time, t	Cash inflow, CF	PVIF _{5%, 5}	CF*PVIF	t*CF*PVIF	Duration, D _{es}
1	\$1,223,912	0.9524	\$1,165,630	\$1,165,630	
2	1,223,912	0.9070	1,110,124	2,220,249	
3	1,223,912	0.8638	1,057,261	3,171,784	
4	1,223,912	0.8227	1,006,915	4,027,662	
5	1,223,912	0.7835	958,967	4,794,835	
			\$5,298,897	\$15,380,160	2.9025
Panel C: Net durations					
Fund raised, B	\$10,000,000		Escrow amount, B _{pf} - B	\$5,298,250	
Face value of bond, B _{pf}	\$15,298,250		Escrow return, r _{es}	5%	
Bond yield, r _{pf}	8.00%		Escrow weight, (B _{pf} - B)/B	0.530	
Bond weight, B _{pf} /B	1.530		Escrow duration, D _{es}	2.903	
Bond duration, D _{pf}	4.312		Net duration, D _{net}	5.059	

If escrow return is less than the bond yield, i.e., $r_{es} = r_{pf}$, then the net duration exceeds the maturity

respectively. The net duration, D_{net} , of the pre-funded issue increases to 5.059 years. The interest subsidy creates the additional leverage and which stretches the duration beyond its maturity.⁴ Because interest subsidy is a realistic condition, the pre-funded bond issue has greater interest rate risk than the comparable zero-coupon bond.

Table 5 presents net duration, interest subsidy and loss of value associated with a pre-funded bond issue for different bond yields and escrow returns. When $r_{pf} = r_{es}$ then there is no interest subsidy or loss of value and the net duration of the pre-funded issue is equal to bond maturity. The net duration, interest subsidy and loss of value increases with the increase in the spread, $r_{pf} = r_{es}$.

Table 6 presents the case when pre-funded bonds are priced to yield lower than the zero-coupons. The asset-based coupon payments of the pre-funded issue are default free, thus market lowers the yield by 25, 50 or 75 basis points from the comparable zero-coupon yield. We recalculate the total face value, net duration, interest subsidy and loss of value under these conditions. Results in Table 6 indicate that impact of the

⁴This is analogous to a situation in portfolio construction. Consider two assets with standard deviations 10% and 20%. For an investor who is long on both assets, the portfolio standard deviation will be between 10% and 20%. However, if the investor is short on first asset and long on the second asset then portfolio standard deviation will exceed 20%.

Table 5 Net duration, interest subsidy and value loss of pre-funded bonds. Pre-tax interest subsidy = $(r_{pf} B_{pf})(r_{pf} - r_{es})$ per year Value Loss = $(r_{pf} B_{pf})(r_{pf} - r_{es}) \frac{(1+r_{pf})^n - 1}{(1+r_{pf})^n}$

	Escrow return, r_{es} (%)	Pre-funded bond yield, r_{pf}				
			8%	7%	6%	5%
Maturity, n 5 years	8	Net duration, years	5			
		Interest subsidy	0			
		Value loss	0			
	7	Net duration, years	5.019	5		
		Interest subsidy	\$11,905	0		
		Value loss	(\$47,533)	0		
	6	Net duration, years	5.038	5.016	5	
		Interest subsidy	\$24,132	\$9,927	0	
		Value loss	(\$96,353)	(\$40,703)	0	
	5	Net duration, years	5.059	5.037	5.013	5
		Interest subsidy	\$36,717	\$20,088	\$8,106	0
		Value loss	(\$146,602)	(\$82,364)	(\$34,144)	0
Maturity, n 10 years	8	Net duration, years	10			
		Interest subsidy	0			
		Value loss	0			
	7	Net duration, years	10.198	10		
		Interest subsidy	\$18,260	0		
		Value loss	(\$122,527)	0		
	6	Net duration, years	10.433	10.165	10	
		Interest subsidy	\$38,911	\$14,439	0	
		Value loss	(\$261,097)	(\$101,414)	0	

(continued)

Table 5 (continued)

	Escrow return, r_{es} (%)	Pre-funded bond yield, r_{pf}				
			8%	7%	6%	5%
	5	Net duration, years	10.715	10.358	10.135	10
		Interest subsidy	\$62,784	\$30,469	\$11,180	0
		Value loss	(\$421,288)	(\$214,004)	(\$82,282)	0

Empty cell represent the improbable case of $r_{pf} < r_z$

Zero-coupon and pre-funded bonds are priced by market as equivalent securities

Table 6 Face value, net duration, interest subsidy and value loss of pre-funded bonds pre-funded bonds are priced to yield lower than comparable zero-coupon bonds

Discount rate on zero-coupon bonds, r_z	Face value of zero-coupon bonds, B_z	Pre-funded bond yield, B_{pf}				
			r_z	$r_z - .25\%$	$r_z - .50\%$	$r_z - .75\%$
		Face value of pre-funded, B_{pf}	\$26,160,123	\$24,902,535	\$23,760,313	\$22,718,277
8%	\$21,589,250	Duration, D_{net}	10.718 years	10.611 years	10.516 years	10.432 years
		Interest subsidy	\$62,784	\$53,074	\$44,551	\$37,059
		Value loss	(\$421,288)	(\$360,179)	(\$305,799)	(\$257,307)
		Face value of Pre-funded, B_{pf}	\$21,763,801	\$20,886,293	\$20,076,805	\$19,327,721
7%	\$19,671,514	Duration, D_{net}	10.358 years	10.291 years	10.266 years	10.181 years
		Interest subsidy	\$30,469	\$24,672	\$19,575	\$15,100
		Value loss	(\$214,004)	(\$175,306)	(\$140,721)	(\$109,831)
		Face value of pre-funded, B_{pf}	\$18,632,525	\$17,985,604	\$17,382,097	\$16,817,777
6%	\$13,382,256	Duration, D_{net}	10.135 years	10.094 years	10.058 years	10.027 years
		Interest subsidy	\$11,180	\$7,756	\$4,780	\$2,207
		Value loss	(\$82,282)	(\$57,769)	(\$36,030)	(\$16,839)

Maturity = 10 years and escrow account yield = 5%

spread, $r_{pf} - r_{es}$ is still dominant. The total face value and net duration of the pre-funded issue is greater than corresponding values for the zero-coupon bond.

4.5 Conclusion

Pre-funded coupon bonds have been developed and sold by investment bankers in place of zero-coupon bonds to raise funds for companies facing cash flow problems. Additional bonds are issued and proceeds are deposited in an escrow account to finance the coupon payment. Our analysis indicates that when the pre-funded bond yield is the same as the escrow return then total face value of the pre-funded issued is the same as the total face value of the zero-coupon bonds and the effective cost of pre-funded issue is the same as the yield to maturity of the zero-coupon bond. Also, increase in the spread between pre-funded bond yield and zero-coupon yield increases the total face value of the bonds issued and its effective cost. The interest subsidy creates additional leverage, which stretches the net duration of the pre-funded issue beyond its maturity. Increase in the spread between pre-funded bond yield and zero-coupon yield increases net duration, interest subsidy and loss of value. Even when pre-funded bonds are priced to yield lower than the zero-coupons, impact of the spread is dominant – total face value and net duration of the pre-funded issue is still greater than corresponding values for the zero-coupon bond.

References

- Bierman, Harold. 2005. Collateralization of bond interest: An investor's perspective. *Journal of Investing* 4: 32–34.
- Brigham, Eugene F., and Phillip R. Daves. 2010. *Intermediate financial management*. Mason: Thomson Southwestern Publishing.
- Doherty, Jacqueline. 1997. For junk borrowers, pre-funded bonds pick up steam, but they may pose greater risk than zeros. *Barrons*, (December), MW15.
- Fabozzi, Frank J. 2010. *Bond markets, analysis and strategies*. Englewood Cliffs, NJ: Prentice Hall.
- Goodman, Laurie S., and Alan H. Cohen, (1989) Payment-in-Kind debentures: An innovation. *Journal of Portfolio Management* 9–19.
- Myers, S.C. 1984. The capital structure puzzle. *Journal of Finance* 39: 575–592.
- Macaulay, Frederick. 1938. Some theoretical problems suggested by the movement of interest rates, bond yields, and stock prices in the U.S. *Since 1856* (National Bureau of Economic Research, New York).
- Redington, F.M. 1952. Review of the principles of life office valuation. *Journal of the Institute of Actuaries* 286–340.
- Ross, Stephen A., Randolph W. Westerfield, and Jeffrey Jaffe. 2010. *Corporate finance*. Homewood: Irwin McGraw Hill.
- Samuelson, Paul A. 1945. The effect of interest rate increases on the banking system. *American Economic Review* 16–27.
- Zeng, Min. 2008. Zero-Coupon bonds see pop in demand. *Wall Street Journal*, New York, N.Y., (December 4), 9.

Intertemporal Risk and Currency Risk

5

Jow-Ran Chang and Mao-Wei Hung

Abstract

Empirical work on portfolio choice and asset pricing has shown that an investor's current asset demand is affected by the possibility of uncertain changes in future investment opportunities. In addition, different countries have different prices for goods when there is a common numeraire in the international portfolio choice and asset pricing. In this survey, we present an intertemporal international asset pricing model (IAPM) that prices market hedging risk and exchange rate hedging risk in addition to market risk and exchange rate risk. This model allows us to explicitly separate hedging against changes in the investment opportunity set from hedging against exchange rate changes as well as separate exchange rate risk from intertemporal hedging risk.

Keywords

Currency risk · Exchange rate risk · Hedging risk · Inflation risk · International asset pricing · Intertemporal asset pricing · Intertemporal risk · Intertemporal substitution · Purchasing power parity · Recursive preference · Risk aversion

5.1 Introduction

In a dynamic economy, it is often believed that if investors anticipate information shifts, they will adjust their portfolios to hedge these shifts. To capture the dynamic hedging effect, Merton (1973) developed a continuous-time asset pricing model

J.-R. Chang (✉)

National Tsing Hua University, Hsinchu, Taiwan, Republic of China

e-mail: jrchang@mx.nthu.edu.tw

M.-W. Hung

National Taiwan University, Taipei, Taiwan, Republic of China

which explicitly takes into account hedging demand. In contrast to the Arbitrage Pricing Theory (APT) framework, there are two factors, which are theoretically derived from Merton's model: a market factor and a hedging factor. Stulz (1981) extended the intertemporal model of Merton (1973) to develop an international asset pricing model. However, an empirical investigation is not easy to implement in the continuous-time model. In a recent paper, Campbell (1993) developed a discrete-time counterpart of Merton's model. Motivated by Campbell's results, Chang and Hung (2000) adopted a conditional two-factor asset pricing model to explain the cross-sectional pricing relationships among international stock markets. In their setup, assets are priced using their covariance with the market portfolio as well as with the hedging portfolio, both of which account for changes in the investment set. Under their proposed international two-factor asset pricing model framework, the international capital asset pricing model (CAPM) is misspecified and estimates of the CAPM model are subject to the omitted variable bias.

If purchasing power parity (PPP) is violated, investors from different countries will have different evaluations for real returns for investment in the same security. This implies that the optimal portfolio choices are different across investors residing in different countries, and any investment in a foreign asset is exposed to currency risk. Therefore, it is reasonable to assume that investors from different countries have different estimations for real returns. This phenomenon clearly shows the existence of currency risk as well as market risk.

There are two goals in this survey. First, we want to know whether hedging demand is important to an international investor. Second, we want to separate currency hedging risk from intertemporal market hedging risk on an international asset pricing model.

The approach we describe here was first proposed by Epstein and Zin (1989, 1991). In their model, the investor is assumed to use a nonexpected utility that distinguishes the coefficient of relative risk aversion and the elasticity of intertemporal substitution. Campbell (1993) applied a log-linear approximation to the budget constraint in order to replace consumption from a standard intertemporal asset pricing model. Chang and Hung (2000) used this model to explain the intertemporal behavior in the international financial markets under no differences in consumption opportunity set.

An important challenge therefore remains – how to build a more realistic intertemporal international asset pricing model (e.g., when the consumption opportunity set is different). This chapter surveys the progress that has been made on this front, drawing primarily from Chang and Hung (2000) and Chang et al. (2005).

In Sect. 5.2, we present a testable intertemporal capital asset pricing model proposed by Campbell. Hence, we can examine whether Campbell's model explains the intertemporal behavior of a number of international financial markets. In Sect. 5.3, we separate currency hedging risk from intertemporal market hedging risk. This is accomplished by extending Campbell's model to an international framework in which investor's utility depends on real returns rather than on nominal returns and PPP deviation.

5.2 No Differences in Consumption Opportunity Set

This section describes the international asset pricing model we employ to estimate and test the pricing relationships among the world's five main equity markets. The model we use is a two-factor model based on Campbell (1993). We first review the theory of nonexpected utility proposed by Weil (1989) and Epstein and Zin (1991). Then we apply a log-linear approximation to the budget constraint to derive an international asset pricing model, which is used in this chapter.

5.2.1 Asset Pricing Model

5.2.1.1 Nonexpected Utility

We consider an economy in which a single, infinitely lived representative international agent chooses consumption and portfolio composition to maximize utility and uses U.S. dollar as the numeraire and where there is one good and N assets in the economy. The international agent in this economy is assumed to be different to the timing of the resolution of uncertainty over temporal lotteries. The agent's preferences are assumed to be represented recursively by

$$V_t = W(C_t, E_t[V_{t+1} | I_t]), \quad (1)$$

where $W(.,.)$ is the aggregator function, C_t is the consumption level at time t , and E_t is the mathematical expectation conditional on the information set at time t . As shown by Kreps and Porteus (1978), the agent prefers early resolution of uncertainty over temporal lotteries if $W(.,.)$ is convex in its second argument. Alternatively, if $W(.,.)$ is concave in its second argument, the agent will prefer late resolution of uncertainty over temporal lotteries.

The aggregator function is further parameterized by

$$\begin{aligned} V_t &= \left[(1 - \delta)C_t^{1-\rho} + \delta(E_t V_{t+1}^{1-\lambda})^{(1-\rho)/(1-\lambda)} \right]^{1/(1-\rho)} \\ &= \left[(1 - \delta)C_t^{(1-\lambda)/\theta} + \delta(E_t V_{t+1}^{1-\lambda})^{1/\theta} \right]^{\theta/(1-\lambda)} \end{aligned} \quad (2)$$

Parameter δ is the agent's subjective time-discount factor and λ is interpreted as the Arrow–Pratt coefficient of relative risk aversion. In addition, $1/\rho$ measures the elasticity of intertemporal substitution. For instance, if the agent's coefficient of relative risk aversion (λ) is greater than the reciprocal of the agent's elasticity of intertemporal substitution (ρ), then the agent would prefer an early resolution towards uncertainty. Conversely, if the reciprocal of the agent's elasticity of intertemporal substitution is larger than the agent's coefficient of relative risk aversion, then the agent would prefer a late resolution of uncertainty. If λ is equal to ρ , the agent's utility becomes an isoelastic, von Neumann–Morgenstern utility, and the agent would be indifferent to the timing of the resolution of uncertainty.

Furthermore, θ is defined as $\theta = (1 - \lambda)/(1 - \rho)$ in accordance with Giovannini and Weil (1989). Three special cases are worth mentioning. First, $\theta \rightarrow 0$ when $\lambda \rightarrow 1$. Second, $\theta \rightarrow \infty$ when $\rho \rightarrow 1$. Third, $\theta = 1$ when $\lambda = \rho$. Under these circumstances, Eq. 2 becomes the von Neumann–Morgenstern expected utility

$$V_t = \left[(1 - \delta) E_t \sum_{j=1}^{\infty} \delta^j \tilde{C}_{t+j}^{1-\gamma} \right]^{1/(1-\lambda)} \quad (3)$$

5.2.1.2 Log-Linear Budget Constraint

We now turn to the characterization of the budget constraint of the representative investor who can invest wealth in N assets. The gross rate of return on asset i held throughout period t is given by $R_{i,t+1}$. Let

$$R_{m,t+1} = \sum_{i=1}^N \alpha_{i,t} R_{i,t+1}, \quad (4)$$

denote the rate of return on the market portfolio, and $\alpha_{i,t}$ be the fraction of the investor's total wealth held in the i th asset in period t . There are only $N - 1$ independent elements in $\alpha_{i,t}$ since the constraint

$$\sum_{i=1}^N \alpha_{i,t} = 1 \quad (5)$$

holds for all t . The representative agent's dynamic budget constraint can be given by

$$W_{t+1} = R_{m,t+1}(W_t - C_t), \quad (6)$$

where W_{t+1} is the investor's wealth at time t . The budget constraint in Eq. 6 is nonlinear because of the interaction between subtraction and multiplication. In addition, the investor is capable of affecting future consumption flows by trading in risky assets. Campbell linearizes the budget constraint by dividing Eq. 6 by W_t , taking log, and then using a first-order Taylor expansion around the mean log consumption/wealth ratio, $\log(C/W)$. If we define the parameter $\beta = 1 - \exp(c_t - w_t)$, the approximation to the intertemporal budget constraint is

$$\Delta w_{t+1} \cong r_{m,t+1} + k + \left(1 - \frac{1}{\beta}\right)(c_t - w_t), \quad (7)$$

where the log form of the variable is indicated by lowercase letters and k is a constant.

Combining Eq. 7 with the following equality,

$$\Delta w_{t+1} = \Delta c_{t+1} + (c_t - w_t) - (c_{t+1} - w_{t+1}), \quad (8)$$

we obtain a different equation in the log consumption/wealth ratio, $c_t - w_t$. Campbell (1993) shows that if the log consumption/wealth ratio is stationary, i.e., $\lim_{j \rightarrow \infty} \beta^j (c_{t+j} - w_{t+j}) = 0$, then the approximation can be written as

$$\begin{aligned} c_{t+1} - E_t c_{t+1} &= (E_{t+1} - E_t) \sum_{j=0}^{\infty} \beta^j r_{m,t+1+j} \\ &\quad - (E_{t+1} - E_t) \sum_{j=1}^{\infty} \beta^j \Delta c_{t+1+j}. \end{aligned} \quad (9)$$

Equation 9 can be used to express the fact that an unexpected increase in consumption today is determined by an unexpected return on wealth today (the first term in the first sum on the right-hand side of the equation), or by news that future returns will be higher (the remaining terms in the first sum), or by a downward revision in expected future consumption growth (the second sum on the right-hand side).

5.2.1.3 Euler Equations

In this setup, Epstein and Zin (1989) derive the following Euler equation for each asset:

$$1 = E_t \left[\left\{ \delta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \right\}^{\theta} \left\{ \frac{1}{R_{m,t+1}} \right\}^{1-\theta} R_{i,t+1} \right] \quad (10)$$

Assume for the present that asset prices and consumption are jointly lognormal or apply a second-order Taylor expansion to the Euler equation. Then, the log version of the Euler equation (10) can be represented as

$$\begin{aligned} 0 &= \theta \log \delta - \theta \rho E_t \Delta c_{t+1} + (\theta - 1) E_t r_{m,t+1} \\ &\quad + E_t r_{i,t+1} + \frac{1}{2} \left[(\theta \rho)^2 V_{cc} + (\theta - 1)^2 V_{mm} \right. \\ &\quad + V_{ii} - 2\theta \rho (\theta - 1) V_{cm} - 2\theta \rho V_{ci} \\ &\quad \left. + 2(\theta - 1) V_{im} \right] \end{aligned} \quad (11)$$

where V_{cc} denotes $\text{var}(c_{t+1})$, V_{jj} denotes $\text{var}(r_{j,t+1}) \forall j = i, m$, V_{cj} denotes $\text{cov}(c_{t+1}, r_{j,t+1}) \forall j = i, m$, and V_{im} denotes $\text{cov}(r_{i,t+1}, r_{m,t+1})$.

By replacing asset i by the market portfolio and rearranging Eq. 11, we obtain a relationship between expected consumption growth and expected return on the market portfolio

$$\begin{aligned} E_t \Delta c_{t+1} &= \frac{1}{\rho} \log \delta + \frac{1}{2} \left[\theta \rho V_{cc} + \left(\frac{\theta}{\rho} \right) V_{mm} \right. \\ &\quad \left. - 2\theta V_{cm} \right] + \frac{1}{\rho} E_t r_{m,t+1}. \end{aligned} \quad (12)$$

When we subtract Eq. 11 for the risk-free asset from that for asset i , we obtain

$$E_t r_{i,t+1} - r_{f,t+1} = -\frac{V_{ii}}{2} + \theta(\rho V_{ic}) + (1 - \theta)V_{im} \quad (13)$$

where $r_{f,t+1}$ is a log riskless interest rate. Equation 13 expresses the expected excess log return on an asset (adjusted for Jensen's inequality effect) as a weighted sum of two terms. The first term, with a weight θ , is the asset covariance with consumption multiplied by the intertemporal elasticity of substitution, ρ . The second term, with a weight $1 - \theta$, is the asset covariance with the return from the market portfolio.

5.2.1.4 Substituting Consumption Out of the Asset Pricing Model

Now, we combine the log-linear Euler equation with the approximated log-linear budget constraint to get an intertemporal asset pricing model without consumption. Substituting Eq. 12 into Eq. 9, we obtain

$$c_{t+1} - E_t c_{t+1} = r_{m,t+1} - E_t r_{m,t+1} + \left(1 - \frac{1}{\rho}\right) (E_{t+1} - E_t) \sum_{j=1}^{\infty} \beta^j r_{m,t+1+j} \quad (14)$$

Equation 14 implies that the unexpected consumption comes from an unexpected return on invested wealth today or expected future returns.

Based on Eq. 14, the conditional covariance of any asset return with consumption can be rewritten in terms of the covariance with the market return and revisions in expectations of future market returns which is given by

$$\text{cov}_t(r_{i,t+1}, \Delta c_{t+1}) \equiv V_{ic} = V_{im} + \left(1 - \frac{1}{\rho}\right) V_{ih} \quad (15)$$

$$\text{where } V_{ih} = \text{cov}_i \left(r_{i,t+1}, (E_{t+1} - E_t) \sum_{j=1}^{\infty} \beta^j r_{m,t+1+j} \right).$$

Substituting Eq. 15 into Eq. 13, we obtain an international asset pricing model that is not related to consumption:

$$E_t r_{i,t+1} - r_{f,t+1} = -\frac{V_{ii}}{2} + \lambda V_{im} + (\lambda - 1) V_{ih}. \quad (16)$$

Equation 16 states that the expected excess log return in an asset, adjusted for Jensen's inequality effect, is a weighted average of two covariances – the covariance with the return from the market portfolio and the covariance with news about future returns on invested wealth.

5.2.2 Empirical Evidence

The relationship between risk and return has been the focus of recent finance research. Numerous papers have derived various versions of the international asset pricing model. For example, Solnik (1974) extends the static Capital Asset Pricing Model of Sharpe (1964) and Lintner (1965) to an international framework. His empirical findings reveal that national factors are important in the pricing of stock markets. Furthermore, Korajczyk and Viallet (1989) propose that the international CAPM outperforms its domestic counterpart in explaining the pricing behavior of equity markets.

In a fruitful attempt to extend the conditional version of the static CAPM, Harvey (1991) employs the Generalized Method of Moments (GMM) to examine an international asset pricing model that captures some of the dynamic behavior of the country returns. De Santis and Gerard (1997) test the conditional CAPM on international stock markets, but they apply a parsimonious Generalized Auto-Regressive Conditional Heteroscedasticity (GARCH) parameterization as the specification for second moments. Their results indicate that a one-factor model cannot fully explain the dynamics of international expected returns and the price of market risk is not significant.

On the other hand, recent studies have applied the APT of Ross (1976) to an international setting. For instance, Cho et al. (1986) employ factor analysis to demonstrate that additional factors other than covariance risk are able to explain the international capital market. Ferson and Harvey (1993) investigate the predictability of national stock market returns and its relation to global economic risk. Their model includes a world market portfolio, exchange rate fluctuations, world interest rates, and international default risk. They use multifactor asset pricing models with time-varying risk premiums to examine the issue of predictability. But, one of the drawbacks of the APT approach is that the number and identity of the factors are determined either ad hoc or statistically from data rather than from asset pricing models directly.

Several international asset pricing models explicitly take into account currency risk, for example, see Solnik (1974), Stulz (1981), and Adler and Dumas (1983). But investors in these models are assumed to maximize a time-additive, von Neumann–Morgenstern expected utility of lifetime consumption function. This implies that two distinct concepts of intertemporal substitution and risk aversion are characterized by the same parameter. Another approach examines consumption risk. Cumby (1990) proposes a consumption-based international asset pricing model. Difficulty occurs in the usage of aggregate consumption data, which are measured with error, and are time-aggregated. Chang and Hung (2000) show that estimations of price of market risk obtained from the De Santis and Gerard (1997) conditional CAPM model may be biased downward due to the omission of the hedging risk, which is negatively correlated to the market risk.

5.3 Differences in Consumption Opportunity Set

In this section, we consider the problem of optimal consumption and portfolio allocation in a unified world capitalmarket with no taxes and transactions costs. Moreover, investors' preferences are assumed to be nationally heterogeneous and

asset selection is the same for investors in different countries. Consider a world of $M + 1$ countries and a set of S equity securities. All returns are measured in the $M + 1$ st country's currency in excess of the risk-free rate and this currency is referred to as the numeraire currency. Investors are assumed to maximize Kreps–Porteus utility for their lifetime consumption function.

5.3.1 Portfolio Choice in an International Setting

5.3.1.1 Kreps–Porteus Preferences

Define C_t as the current nominal consumption level at time t , and P_t as the price level index at time t , expressed in the numeraire currency. In the setup of Kreps and Porteus (1978) nonexpected utility, the investor's value function can be represented as:

$$V_t = U \left[\frac{C_t}{P_t}, E_t V_{t+1} \right]. \quad (17)$$

where V_t is the lifetime utility at time t , V_{ej} is the expected value function conditional on the information available to the investor at time t , $U[.,.]$ is the aggregator function that aggregates current consumption with expected future value. As shown by Kreps and Porteus, the agent prefers early resolution of uncertainty over temporal lotteries if $U[.,.]$ is convex in its second argument. On the other hand, if $U[.,.]$ is concave in its second argument, the agent will prefer late resolution of uncertainty over temporal lotteries.

Furthermore, the aggregator function is parameterized to be homogenous of degree one in current real consumption and in the value of future state-dependent real consumption:

$$U \left[\frac{C_t}{P_t}, E_t V_{t+1} \right] = \left[(1 - \delta) \left(\frac{C_t}{P_t} \right)^{1-\rho} + \delta (E_t V_{t+1})^{(1-\rho)/(1-\lambda)} \right]^{(1-\lambda)/(1-\rho)}, \quad (18)$$

where λ is the Arrow–Pratt coefficient of relative risk aversion, ρ can be interpreted as the elasticity of intertemporal substitution, and $\delta \in (0, 1)$ is the subjective discount factor.

The Kreps–Porteus preference allows the separation of risk aversion from intertemporal substitution. For instance, if the agent's coefficient of relative risk aversion, λ , is greater than the reciprocal of the agent's elasticity of intertemporal substitution, ρ , then the agent prefers early resolution of uncertainty. Conversely, if the reciprocal of the agent's elasticity of intertemporal substitution is larger than the agent's coefficient of relative risk aversion, the agent prefers late resolution of uncertainty. When $\rho = \lambda$, the objective function is the time-separable power utility function with relative risk aversion λ . In addition, when both λ and ρ equal 1, we have standard time-separable log utility function. Hence, the standard time- and state-separable expected utility is a special case under Kreps–Porteus preferences.

5.3.1.2 Optimal Consumption and Portfolio Allocation

We now turn to the characterization of the budget constraint of the representative investor who can invest his wealth in $N(= M + S)$ assets that include M currencies and S equities. Currencies may be taken as the nominal bank deposits denominated in the nonnumeraire currencies. The gross rate of nominal return on asset i held throughout period t is given by $R_{i,t+1}$. Let

$$R_{m,t+1} \equiv \sum_{i=1}^N \alpha_{i,t} R_{i,t+1} \quad (19)$$

denote the rate of return on the market portfolio, and $\alpha_{i,t}$ be the fraction of the investor's total wealth held in the i th asset in period t . There are only $N - 1$ independent elements in $\alpha_{i,t}$, since the constraint

$$\sum_{i=1}^N \alpha_{i,t} = 1 \quad (20)$$

holds for all t . The representative agent's dynamic budget constraint in terms of real variables can be written as:

$$\frac{W_{t+1}}{P_{t+1}} = R_{m,t+1} \frac{P_t}{P_{t+1}} \left(\frac{W_t}{P_t} - \frac{C_t}{P_t} \right) \quad (21)$$

where W_{t+1} is the investor's nominal wealth at time t . The budget constraint in Eq. 21 is nonlinear because of the interaction between subtraction and multiplication.

Define I_t as the information set available to the representative agent at time t . Denoting by $V(W/P, I)$ the maximum value of Eq. 17 subject to Eq. 20, the standard Bellman equations can then be written as:

$$\begin{aligned} V\left(\frac{W_t}{P_t}, I_t\right) = \max_{C_t, \{\alpha_{i,t}\}_{i=1}^N} & \left\{ (1 - \delta) \left(\frac{C_t}{P_t} \right)^{1-\rho} \right. \\ & \left. + \delta \left[E_t V\left(\frac{W_{t+1}}{P_{t+1}}, I_{t+1}\right) \right]^{(1-\rho)/(1-\lambda)} \right\}^{(1-\lambda)/(1-\rho)} \end{aligned} \quad (22)$$

Due to the homogeneity of the recursive structure of preferences, the value function can be written in the following functional form:

$$V\left(\frac{W_t}{P_t}, I_t\right) = \Phi(I_t) \left(\frac{W_t}{P_t} \right)^{t-\lambda} \equiv \Phi_t \left(\frac{W_t}{P_t} \right)^{1-\lambda}, \quad (23)$$

where $\Phi(\cdot)$ is an unknown function. The homogeneity of degree zero of the recursive utility function implies that $V(W/P, I)$ satisfying Eq. 22 must be homogeneous of degree zero in W and P .

Let the derivatives with respect to the decision variables C_t equal zero, we then obtain:

$$C_t^{-\rho} = \frac{\delta}{1-\delta} \psi(W_t - C_t)^{-\rho} \quad (24)$$

$$\text{where } \psi_t = E_t \left[\Phi_{t+1} \left(R_{m,t+1} \frac{P_t}{P_{t+1}} \right)^{1-\lambda} \right]^{(1-\rho)/(1-\lambda)}$$

Given the structure of the problem, the nominal consumption function is linear in nominal wealth. Hence, we can rewrite Eq. 24 as:

$$\mu_t^{-\rho} = \frac{\delta}{1-\delta} \psi(1 - \mu_t)^{-\beta} \quad (25)$$

where $C(W_t, I_t) = \mu(I_t)W_t \equiv \mu_t W_t$. Combining the Envelope condition with respect to W_t with the first-order condition in Eq. 25, we obtain the following functional form:

$$\Phi_t = (1 - \delta)^{(1-\lambda)/(1-\rho)} \left[\left(\frac{C_t}{W_t} \right)^{-\rho} \right]^{(1-\lambda)/(1-\rho)} \quad (26)$$

Substituting this expression into Eq. 25, we obtain the following Euler equation for optimal consumption decision:

$$E_t \left\{ \left[\delta \left(\frac{C_{t+1}/P_{t+1}}{C_t/P_t} \right)^{-\rho} \right]^{(1-\lambda)/(1-\rho)} \left(R_{m,t+1} \frac{P_t}{P_{t+1}} \right)^{(1-\lambda)/(1-\rho)} \right\} = 1, i = 1, \dots, N \quad (27)$$

The maximization with respect to the decision variable $\alpha_i (i = 2, \dots, N)$, given $\alpha_1 = 1 - \sum_{i=2}^N \alpha_i$, on the right hand side of Eq. 22, is equivalent to the following problem:

$$\begin{aligned} \max_{\{\alpha_{i,t}\}_{i=2}^N} E_t & \left[\Phi_{t+1} \left(\sum_{i=1}^N \alpha_{i,t} R_{i,t+1} \frac{P_t}{P_{t+1}} \right)^{1-\lambda} \right] \\ \text{s.t. } & \sum_{i=1}^N \alpha_{i,t} = 1 \end{aligned} \quad (28)$$

Using this optimal problem along with Eq. 26, it is straightforward to show that the necessary conditions can be derived as:

$$\begin{aligned} E_t & \left\{ \left[\delta \left(\frac{C_{t+1}/P_{t+1}}{C_t/P_t} \right)^{-\rho} \right]^{1-\lambda/1-\rho} \left(R_{m,t+1} \frac{P_t}{P_{t+1}} \right)^{[(1-\lambda)/(1-\rho)]-1} \right. \\ & \left. (R_{i,t+1} - R_{1,t+1}) \frac{P_t}{P_{t+1}} \right\} = 0, i = 1, \dots, N \end{aligned} \quad (29)$$

Taking Eqs. 27 and 29 together to represent the Euler equations of the optimal problem defined in Eq. 22, we obtain a set of N equations that provide a more direct comparison with the traditional expected utility Euler equations. Multiplying Eq. 29 by $\alpha_{i,t}$, summing up by i , and substituting from Eq. 27, we obtain:

$$E_t \left\{ \left[\delta \left(\frac{C_{t+1}/P_{t+1}}{C_t/P_t} \right)^{-\rho} \right]^\theta \left(R_{m,t+1} \frac{P_t}{P_{t+1}} \right)^{\theta-1} R_{i,t+1} \frac{P_t}{P_{t+1}} \right\} = 1, i = 1, \dots, N \quad (30)$$

where $\theta = (1 - \lambda)/(1 - \rho)$. These are the real form Euler equations which are similar to the nominal form Euler equations seen in Epstein and Zin (1989).

When $\rho = \lambda$, the Euler equations of the time additive expected utility model are also obtained in terms of real variables:

$$E_t \left[\delta \left(\frac{C_{t+1}/P_{t+1}}{C_t/P_t} \right)^{-\rho} R_{i,t+1} \frac{P_t}{P_{t+1}} \right] = 1, i = 1, \dots, N \quad (31)$$

Another special case of this model is the logarithmic risk preferences where $\rho = \lambda = 1$. Then, the real Euler equations are equal to the nominal Euler equations, and can be written in two algebraically identical functional forms:

$$V_i^{\text{cr}} \quad (32)$$

or

$$E_t[R_{i,t+1}/R_{m,t+1}] = 1, i = 1, \dots, N \quad (33)$$

In this case, the parameter ρ governing intertemporal substitutability cannot be identified from these equations. Hence, there is no difference between Euler equations of the nonexpected utility model with logarithmic risk preferences and those of the expected utility model with logarithmic risk preferences.

Assume that asset prices and consumption are jointly lognormal or use a second-order Taylor expansion in the Euler equation when we assume that asset prices and consumption are conditional homoskedastic, then the log-version of the real Euler equation (30) can be represented as:

$$\begin{aligned} \theta &= \theta \log \delta - \theta \rho E_t \Delta c_{t+1} + (\theta - 1) E_t r_{m,t+1} \\ &\quad + E_t r_{i,t+1} + \theta(\rho - 1) E_t \Delta \pi_{t+1} \\ &\quad + \frac{1}{2} \left[(\theta \rho)^2 V_{cc} + (\theta - 1)^2 V_{mm} + V_{ii} \right. \\ &\quad \left. - 2\theta \rho (\theta - 1) V_{cm} - 2\theta \rho V_{ci} + 2(\theta - 1) V_{im} \right] \\ &\quad + \frac{1}{2} \left[(\theta(\rho - 1))^2 V_{\pi\pi} - 2\theta^2 \rho (\rho - 1) V_{\pi c} \right. \\ &\quad \left. + 2\theta(\theta - 1)(\rho - 1) V_{\pi m} + 2\theta(\rho - 1) V_{\pi i} \right] \end{aligned} \quad (34)$$

where V_{cc} denotes $\text{var}_t(c_{t+1})$, V_{jj} denotes $\text{var}_t(r_{j,t+1}) \forall j = i, m$, V_{cj} denotes $\text{cov}_t(c_{t+1}, r_{j,t+1}) \forall j = i, m$, V_{im} denotes $\text{cov}_t(r_{i,t+1}, r_{m,t+1})$, $V_{i\pi} = \text{cov}_t(r_{i,t+1}, \pi_{t+1})$, and $\pi_{t+1} = d \ln(P_{t+1}) = \frac{dP_{t+1}}{P_{t+1}}$.

Replacing asset i by market portfolio and undergoing some rearrangement, we are able to obtain a relationship between expected consumption growth and the expected return on the market portfolio:

$$E_t \Delta c_{t+1} = \mu_m + \frac{1}{\rho} E_t r_{m,t+1} + \left(1 - \frac{1}{\rho}\right) E_t \pi_{t+1} \quad (35)$$

$$\begin{aligned} \text{where } \mu_m = & \frac{1}{\rho} \log \delta + \frac{1}{2} \left[\theta \rho V_{cc} + \theta \frac{1}{\rho} V_{mm} + 2 \left(1 - \frac{1}{\rho}\right) \theta (\rho - 1) V_{\pi\pi} \right] \\ & - \frac{1}{2} \left[2\theta V_{cm} + 2\theta(\rho - 1) V_{\pi c} - 2\theta \left(1 - \frac{1}{\rho}\right) V_{\pi m} \right] \end{aligned}$$

When the second moments are conditional homoskedastic, Eq. 35 indicates that the consumption growth is linearly related to the expected world market return and expected inflation. In addition, the coefficients of these two variables are summed up to 1.

When we subtract the risk free version of Eq. 34 from the general version, we obtain:

$$E_t r_{i,t+1} - r_{f,t+1} = -\frac{V_{ii}}{2} + \theta \rho V_{ic} + (\theta - \theta \rho) V_{i\pi} + (1 - \theta) V_{im} \quad (36)$$

where $r_{f,t+1}$ is a log riskless real interest rate. This result is similar to that of Campbell (1993) except for the inflation term. Equation 36 shows that the expected excess log return on an asset is a linear combination of its own variance, which is produced by Jensen's inequality, and by a weighted average of three covariances. The weights on the consumption, inflation, and market are $\theta \rho$, $(\theta - \theta \rho)$, and $(1 - \theta)$, respectively. Moreover, the weights are summed up to 1. This is one of the most important differences between Campbell's model and our real model.

If the objective function is a time-separable power utility function, a real functional form of a log-linear version of the consumption CAPM pricing formula can thus be obtained:

$$E_t r_{i,t+1} - r_{f,t+1} = -\frac{V_{ii}}{2} + \rho V_{ic} + (1 - \rho) V_{i\pi} \quad (37)$$

The weights on the consumption and inflation are ρ and $(1 - \rho)$, respectively. These weights are also summed up to 1. However, when the coefficient of relative risk aversion $\lambda = 1$, then $\theta = 0$. The model is reduced to the real functional form of

log-linear static CAPM, which is the same as the nominal structure of log-linear static CAPM.

5.3.2 International Asset Pricing Model Without Consumption

In order to get a pricing formula without consumption, we apply the technique of Campbell (1993). Campbell (1993) suggests to linearize the budget constraint by dividing the nominal form of Eq. 21 by W_t , taking log, and then using a first-order Taylor approximation around the mean log consumption/wealth ratio ($\log(C/W)$). Following his approach, approximation of the nominal budget constraint is:

$$\Delta w_{t+1} \cong r_{m,t+1} + \kappa + \left(1 - \frac{1}{\beta}\right)(c_t - w_t) \quad (38)$$

where the log form of the variable is indicated by lowercase letters, $\beta = 1 - \exp(c_t - w_t)$, and κ is a constant.

Combining Eq. 38 with the following trivial equality

$$\Delta w_{t+1} = \Delta c_{t+1} + (c_t - w_t) - (c_{t+1} - w_{t+1}), \quad (39)$$

we obtain a difference equation in the log consumption/wealth ratio, $c_t - w_t$. When the log consumption/wealth ratio is stationary, i.e., $\lim_{j \rightarrow \infty} \beta^j (c_{t+j} - w_{t+j}) = 0$, Eq. 38 implies that the innovation in logarithm of consumption can be represented as the innovation in the discounted present value of the world market return minus the innovation in the discounted present value of consumption growth:

$$\begin{aligned} c_{t+1} - E_t c_{t+1} &= (E_{t+1} - E_t) \sum_{j=0}^{\infty} \beta^j r_{m,t+1+j} \\ &\quad - (E_{t+1} - E_t) \sum_{j=1}^{\infty} \beta^j \Delta c_{t+1+j} \end{aligned} \quad (40)$$

Now we are ready to derive an international asset pricing model without consumption in terms of real variables by connecting the log-linear Euler equation to the approximation log-linear budget constraint. Substituting Eq. 35 into Eq. 40, we obtain:

$$\begin{aligned} c_{t+1} - E_t c_{t+1} &= r_{m,t+1} - E_t r_{m,t+1} \\ &\quad + \left(1 - \frac{1}{\rho}\right)(E_{t+1} - E_t) \sum_{j=1}^{\infty} \beta^j r_{m,t+1+j} \\ &\quad - \left(1 - \frac{1}{\rho}\right)(E_{t+1} - E_t) \sum_{j=1}^{\infty} \beta^j \pi_{m,t+1+j} \end{aligned} \quad (41)$$

Equation 41 implies that an unexpected consumption may come from three sources. The first one is the unexpected return on invested wealth today. The second one is the expected future nominal returns. The direction of influence depends on whether $1/\rho$ is less or greater than 1. When $1/\rho$ is less than 1, an increase (or decrease) in the expected future nominal return increases (or decreases) the unexpected consumption. Conversely, when $1/\rho$ is greater than 1, an increase (or decrease) in the expected future nominal return decreases (or increases) the unexpected consumption. The third one is the inflation in the investor's own country. The direction of influence also depends on whether $1/\rho$ is less or greater than 1. When $1/\rho$ is less than 1, an increase (or decrease) in the inflation decreases (or increases) the unexpected consumption. Conversely, when $1/\rho$ is greater than 1, an increase (or decrease) in the inflation increases (or decreases) the unexpected inflation.

Based on Eq. 41, the conditional covariance of any asset return with consumption can be rewritten in terms of covariance with market return and revisions in expectations of future market return as:

$$\text{cov}_t(r_{i,t+1}, \Delta c_{t+1}) \equiv V_{ic} = V_{im} + \left(1 - \frac{1}{\rho}\right) \quad (42)$$

where

$$V_{ih} = \text{Cov}_t\left(r_{i,t+1}, (E_{t+1} - E_t) \sum_{j=1}^{\infty} \beta^j r_{m,t+1+j}\right)$$

and

$$V_{ih\pi} = \text{Cov}_t\left(r_{i,t+1}, (E_{t+1} - E_t) \sum_{j=1}^{\infty} \beta^j \pi_{t+1+j}\right)$$

Substituting Eq. 42 into Eq. 36, we thus obtain an international asset pricing model, which is not related to consumption:

$$E_t r_{i,t+1} - r_{f,t+1} = -\frac{V_{ii}}{2} + \lambda V_{im} + (\lambda - 1)V_{ih} + (1 - \lambda)V_{i\pi} + (1 - \lambda)V_{ih\pi} \quad (43)$$

The only preference parameter that enters Eq. 43 is the coefficient of relative risk aversion (λ). The elasticity of intertemporal substitution ρ is not present under this international pricing model. Equation 43 states that the expected excess log return in an asset, adjusted for Jensen's inequality effect, is a weighted average of four covariances. These are the covariance with the market return, the covariance with news about future returns on invested wealth, the covariance with return from inflation, and the covariance with news about future inflation. This result is different from both the international model of Adler and Dumas (1983) and the intertemporal

model of Campbell (1993). Adler and Dumas use von Neumann–Morgenstern utility and assume a constant investment opportunity set to derive the international model, and therefore neither V_{ih} or $V_{ih\pi}$ is included in their pricing formula. Since the intertemporal model of Campbell is a domestic model, it does not deal with the issues of inflation and currency that are emphasized in our international asset pricing model.

5.3.3 International Asset Pricing Model When PPP Deviate

Let us now turn to the problem of aggregation across investors. It is true that different investors use different information set and different methods to forecast future world market return and inflation. To obtain the aggregation results, we first superimpose Eq. 43 by a superscript l to indicate optimal condition for an investor l :

$$E_t r_{i,t+1} - r_{f,t+1} = -\frac{V_{ii}}{2} + \lambda^l V_{im}^l + (\lambda^l - 1) V_{ih}^l + (1 - \lambda^l) V_{i\pi}^l + (1 - \lambda^l) V_{ih\pi}^l \quad (44)$$

Then, Eq. 44 can be aggregated across all investors in all countries.

The operation is to multiply Eq. 44 by η^l , which indicates risk tolerance where $\eta^l = 1/\lambda^l$ and to take an average of all investors, where weights are their relative wealth. After aggregating all investors, we obtain:

$$\begin{aligned} E_t r_{i,t+1} - r_{f,t+1} = & -\frac{V_{ii}}{2} + \frac{1}{\eta^m} V_{im}^m + \left(\frac{1}{\eta^m} - 1 \right) \sum_l \omega^l V_{ih}^l \\ & + \left(1 - \frac{1}{\eta^m} \right) \sum_l \omega^l V_{i\pi}^l + \left(1 - \frac{1}{\eta^m} \right) \sum_l \omega^l V_{ih\pi}^l \end{aligned} \quad (45)$$

$$\text{where } \eta^m = \frac{\left(\sum_l W^l \eta^l \right)}{\left(\sum_l W^l \right)} \text{ and } \omega^l = \frac{(1 - \lambda^l) W^l}{\sum_l (1 - \lambda^l) W^l}.$$

There are several interesting and intuitive results in this equation. First, Eq. 45 shows that an international asset risk premium adjusted for one-half its own variance is related to its covariance with four variables. These are the world market portfolio, aggregate of the innovation in discounted expected future world market returns from different investors across countries, aggregate of the inflation from different countries, and aggregate of the innovation in discounted expected future inflation from

different investors across countries. The weights are $1/\eta^m$, $1/\eta^m - 1$, $1 - 1/\eta^m$, and $1 - 1/\eta^m$, respectively. The sum of these weights is equal to 1. Moreover, it is noted that the market hedging risk is a weighted average of world market portfolio for investors from different countries. This is different from the domestic counterpart of Campbell (1993).

Second, an international asset can be priced without referring to its covariance with consumption growth. Rather, it depends on its covariance with world market return, the weighted average of news about future world market return for investors from different countries, inflation, and the weighted average of news about future inflation for investors from different countries.

Third, the coefficient of risk tolerance, η_m , is the only preference parameter that enters Eq. 45. When consumption is substituted out in this model, the coefficient of intertemporal substitution ρ disappears. Similar results have been documented by Kocherlakota (1990) and Svensson (1989). They show that when asset returns are independently and identically distributed over time, the coefficient of intertemporal substitution is irrelevant for asset returns.

If we are willing to make some more assumptions, we can obtain a more compact result. Namely, if investors are assumed to have the same world market portfolio and use the same method to forecast world market portfolio return, we can multiply Eq. 44 by λ^l , and take an average of all investors, where weights are their relative wealth, to get a simple version of the international asset pricing model:

$$E_t r_{i,t+1} - r_{f,t+1} = -\frac{V_{ii}}{2} + \lambda^m V_{im} + (\lambda^m - 1)V_{ih} + (1 - \lambda^m) \sum_l \omega^l V_{i\pi}^l + (1 - \lambda^m) \sum_l \omega^l V_{ih\pi}^l \quad (46)$$

where $\lambda^m = (\sum_l W^l \lambda^l) / (\sum_l W^l)$ and $\omega^l = \frac{(1 - \lambda^l) W^l}{\sum_l (1 - \lambda^l) W^l}$

Both hedging risk V_{ih} and currency risk $V_{i\pi}^l$ are related to expected return. In addition, they all depend on whether λ^m is different from 1 or not.

Furthermore, when we assume that domestic inflation is nonstochastic, the only random component in π would be the relative change in the exchange rate between the numeraire currency and the currency of the country, where the investor resides. Hence, $V_{i\pi}^l$ is a pure measure of the exposure of asset i to the currency risk of the country, where investor l resides and $V_{ih\pi}^l$ is also a measure of the exposure of asset i to hedge against the currency risk of the country, where investor l resides.

Equation 46 also states that the currency risk is different from the hedging risk. However, if V_{ih} and $V_{i\pi}^l$ are large enough, then whether V_{ih} and $V_{i\pi}^l$ and related to expected return depends on whether or not λ^m is different from 1: This may be the reason why Dumas and Solnik (1995) argue that exchange rate risk premium may be equivalent to intertemporal risk premium. But, their conjecture is based on an empirical “horse race” test between international model and intertemporal model rather than a theoretical derivation.

5.4 Conclusion

The international asset pricing model without consumption developed by Chang and Hung (2004) argues that the real expected asset return is determined by market risk, market hedging risk, currency risk, and currency hedging risk. The weights are related only to relative risk aversion. Moreover, the weights are summed up to 1. Their results may be contrasted with the pioneering work of Adler and Dumas (1983), who assume a constant investment opportunity set, thus their model lacks market hedging risk and currency hedging risk.

In the Chang et al. (2004) model, the price of market hedging risk is equal to the negative price of the currency risk. This may be the reason why Dumas and Solnik (1995) argue that currency risk is equivalent to market hedging risk. But, their conjecture is based on a “horse race” test between international model and intertemporal model rather than on a theoretical derivation.

References

- Adler, M., and B. Dumas. 1983. International portfolio selection and corporation finance: A synthesis. *Journal of Finance* 38: 925–984.
- Chang, J.R., and M.W. Hung. 2000. An International asset pricing model with time-varying hedging risk. *Review of Quantitative Finance and Accounting* 15: 235–257.
- Chang, J.R., V. Errunza, K. Hogan, and M.W. Hung. 2005. An intertemporal international asset pricing model: Theory and empirical evidence. *European Financial Management* 11: 173–194.
- Cho, C.D., C.S. Eun, and L.W. Senbet. 1986. International arbitrage pricing theory: An empirical investigation. *Journal of Finance* 41: 313–330.
- Campbell, J.Y. 1993. Intertemporal asset pricing without consumption data. *American Economic Review* 83: 487–512.
- Cumby, R.E. 1990. Consumption risk and international equity returns: some empirical evidence. *Journal of International Money and Finance* 9: 182–192.
- De Santis, G., and B. Gerard. 1997. International asset pricing and portfolio diversification with time-varying risk. *Journal of Finance* 52: 1881–1912.
- Dumas, B., and B. Solnik. 1995. The world price of foreign exchange risk. *Journal of Finance* 50: 445–479.
- Epstein, L.G., and S.E. Zin. 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica* 57: 937–969.
- . 1991. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: an empirical analysis. *Journal of Political Economy* 99: 263–286.
- Ferson, W.E., and R.C. Harvey. 1993. The risk and predictability of international equity returns. *Review of Financial Studies* 6: 527–567.
- Giovannini, A., and P. Weil. 1989. Risk aversion and intertemporal substitution in the capital asset pricing model. *National Bureau of Economic Research* 2824.
- Harvey, R.C. 1991. The world price of covariance risk. *Journal of Finance* 46: 111–157.
- Kocherlakota, N. 1990. Disentangling the coefficient of relative risk aversion from the elasticity of intertemporal substitution: an irrelevance result. *Journal of Finance* 45: 175–190.
- Korajczyk, R.A., and C. Viallet. 1989. An empirical investigation of international asset pricing. *Review of Financial Studies* 2: 553–585.
- Kreps, D., and E. Porteus. 1978. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46: 185–200.

- Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–37.
- Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867–887.
- Ross, S.A. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13: 341–360.
- Sharpe, W. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- Solnik, B. 1974. An equilibrium model of the international capital market. *Journal of Economic Theory* 8: 500–524.
- Stulz, R.M. 1981. A model of international asset pricing. *Journal of Financial Economics* 9: 383–406.
- Svensson, L.E.O. 1989. Portfolio choice with non-expected utility in continuous time. *Economics Letters* 30: 313–317.
- Weil, P. 1989. The equity premium puzzle and the risk free rate puzzle. *Journal of Monetary Economics* 24: 401–421.

Ren-Raw Chen and Jing-Zhi Huang

Abstract

Credit derivatives are instruments used to measure, manage, and transfer credit risk. Recently, there has been an explosive growth in the use of these instruments in the financial markets. This chapter reviews the structure and use of some credit derivative instruments that are popular in practice.

Keywords

Asset swaps · Basket default swaps · Credit default swaps · Credit derivatives · Credit risk · Credit spread options · Credit spreads · Default risk · Default swaps · Total return swaps

6.1 Introduction

Recent years have seen a dramatic expansion in the use of credit derivatives in the financial industry. Credit derivatives are used in the diversification and transfer of credit risk, the ability to leverage, and the creation of new asset classes providing yield enhancement. This growth is likely to continue as institutional investors, broker-dealers, hedge funds, and insurance companies all realize the advantages that these instruments have over the traditional alternatives. In the following, we present an overview of the main credit derivatives such as default swaps, total return swaps, and credit spread options.

R.-R. Chen (✉)
Fordham University, New York, NY, USA
e-mail: rchen@fordham.edu

J.-Z. Huang
Penn State University, Philadelphia, PA, USA

6.2 Asset Swaps

The most basic building block in the credit world is perhaps the asset swap. An asset swap is a simple structure that enables a counterparty receiving fixed payments on a security to exchange the fixed coupon for a floating rate payment at a fixed spread to London Interbank Offered Rate (LIBOR). Historically, banks have used asset swaps to match their long-term fixed-rate assets with their short-term liabilities, i.e., mortgage loans against depositor accounts. In a par asset swap, one party delivers a risky asset to the other in return for par. They then receive the cash flows of a risky bond in return for regular payments of LIBOR plus a fixed spread (or minus a fixed spread if the asset is better quality than LIBOR, e.g., a U.S. Treasury security). The mechanics of this structure are shown in Fig. 1.

This fixed spread is known as the Asset Swap Spread. The key point concerning asset swaps is that the fixed coupons being paid are *effectively* guaranteed by the counterparty even if the underlying risky asset defaults. As a result, the payer of the fixed coupon has a credit exposure to the issuer of the defaulting bond. The asset swap spread is therefore the additional return required by the payer of the fixed coupon to compensate for the credit risk incurred and to repay any difference in price if the bond is trading away from par. The par amount paid up front can be used to purchase a par floater. The overall result for counterparty A has been to take fixed

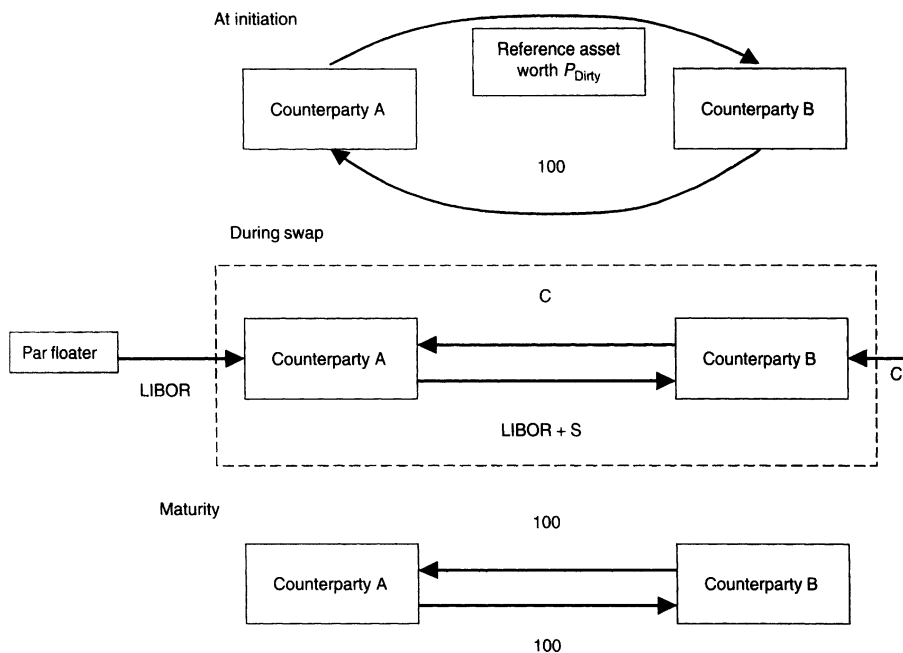


Fig. 1 Mechanics of a par floater asset swap

cash flows from a risky asset and exchange them for the same cash flows paid by a LIBOR quality counterparty. These fixed coupons can then be exchanged for floating rate payments in another standard interest rate swap.

The asset swap is an important measure of default risk as it represents the additional spread that can be locked-in by taking on the risk of an issuer in a fixed-for-floating rate par swap.

6.3 Default Swaps

A default swap is a bilateral contract that allows an investor to buy protection against the risk of default of a specified *reference credit*. The fee may be paid up front, but more often is paid in a “swapped” form as a regular, accruing cash flow.

As a default swap is a negotiated contract, there are several important features that need to be agreed between the counterparties and clearly defined in the contract documentation. First and foremost is the definition of the credit event itself. This is obviously closely linked to the choice of the reference credit and will include such events as bankruptcy, insolvency, receivership, restructuring of debt, and a material change in the credit spread. This last *materiality* clause ensures that the triggering event has indeed affected the price of the reference asset. It is generally defined in spread terms since a fall in the price of the reference asset could also be due to an increase in the level of interest rates.

Many default swaps define the triggering of a credit event using a *reference asset*. However, in many cases, the importance of the reference asset is secondary as the credit event may also be defined with respect to a class of debt issued by a *reference entity*. In this case, the importance of the reference asset arises solely from its use in the determination of the *recovery price* used to calculate the payment following the credit event.

The contract must specify what happens if the credit event occurs. Typically, the protection buyer will usually agree to do one of the following:

- Deliver the defaulted security to the protection seller in return for par in cash. Note that the contract usually specifies a basket of securities that are ranked *pari passu*, which may be delivered in place of the reference asset. In effect, the protection seller is long a “cheapest to deliver” option.
- Receive par minus the default price of the reference asset settled in cash. The price of the defaulted asset is typically determined via a dealer poll conducted within a few weeks to months of the credit event, the purpose of the delay being to let the recovery price stabilize.

These choices are shown in Fig. 2. It is often in the interest of the protection seller to choose physical delivery of the asset since the seller may have the view that either by waiting or by entering into negotiations with the issuer of the reference asset, he may be able to receive more than the default price.

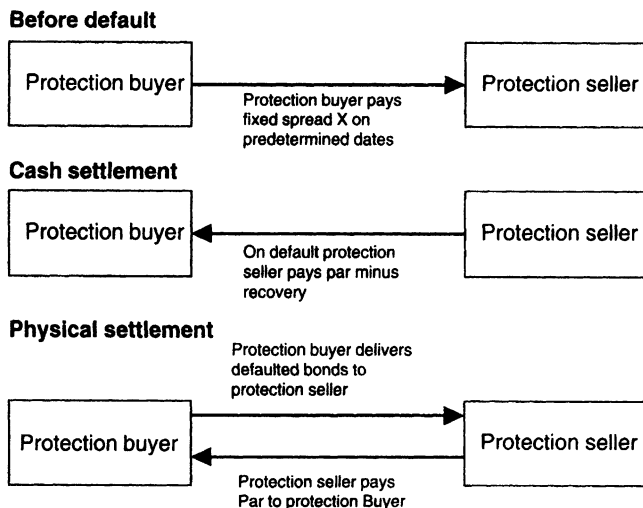


Fig. 2 Mechanics of a default swap

For those familiar with option terminology, it may help to think of a default swap as a knock-in option contingent on the credit event. Until the credit event occurs, the default swap is always out-of-the-money. Even a large deterioration in the credit of the reference asset of a default swap that just stops short of the credit event will not be covered by the default swap.

Some default swaps have a different payoff from the standard par minus recovery price. The main alternative is to have a fixed pre-determined amount that is paid out immediately after the credit event. This is known as a binary default swap. In other cases, where the reference asset is trading at a significant premium or discount to par, the payoff may be tailored to be the difference between the initial price of the reference asset and the recovery price.

The protection buyer automatically stops paying the premium once the credit event has occurred, and this property has to be factored into the cost of the swap payments. It has the benefit of enabling both parties to close out their positions soon after the credit event and so eliminates the ongoing administrative costs that would otherwise occur.

A default swap can be viewed as a form of insurance with one important advantage – efficiency. Provided the credit event in the default swap documentation is defined clearly, the payment due from the triggering of the credit event will be made quickly. Contrast this with the potentially long and drawn out process of investigation and negotiation that may occur with traditional insurance.

In approximate order of importance, the main factors that will determine the cost of the default swap are the shape of the reference asset credit spread curve, the maturity of the protection, the default price of the reference asset, the shape of the LIBOR curve, the credit worthiness of the protection seller, and the correlation of

Table 1 Static replication of a default swap on a par floater that can default on coupon dates only

Event	Long riskless FRN	Short risky FRN via asset swap	Default swap
Coupon payments	+LIBOR	−(LIBOR + ASS)	−ASS
Credit event occurs on coupon date	+100 at Default	−R at default	(100 − R) at default
No credit event	+100 at maturity	−100 at maturity	0

the credit worthiness of the protection seller to the reference asset; default protection bought on the debt of a bank from another closely related bank is probably worthless.

However, it is possible to get a very good idea of the price of the default swap using a simple “static replication” argument. This involves recognizing that buying a default swap on a risky par floating rate asset that only defaults on coupon dates is exactly equivalent to going along a default-free floating rate note and short a risky floating rate note of the same credit quality (see Table 1). If no default occurs, the holder of the position makes a net payment equal to the asset swap spread of the asset on each coupon date until maturity. This spread represents the credit quality of the risky floater at issuance. If default does occur, and we assume that it can only occur on coupon payment dates, the position can be closed out by buying back the defaulted asset in return for the recovery rate and selling the par floater. The net value of the position is equal to the payoff from the default swap.

This argument implies that we would expect the price of the default swap on a par floater to be close to the asset swap spread with the credit quality of the reference credit and the same maturity as the default swap. It also shows that par floaters are a perfect hedge for default swaps.

6.4 Total Return Swaps

A total return swap is a contract that allows an investor to receive all of the cash flow benefits of owning a reference asset without actually possessing the asset itself. The mechanics of this structure are shown in Fig. 3.

At trade inception, one party, the total return receiver, agrees to make payments of LIBOR plus a fixed spread to the other party, the total return payer, in return for the coupons paid by some specified asset. At the end of the term of the total return swap, the total return receiver must then pay the difference between the final market price of the asset and the initial price of the asset. If default occurs, this means that the total return receiver must shoulder the loss by paying the difference between the initial value of the reference asset and the default value of the reference asset. Standard practice is for cash settlement.

It is important to understand that a total return swap has a default swap embedded within it – the payer of the total return has insured himself against the default risk of the asset. Furthermore, he is also protected from the price risk of the asset.

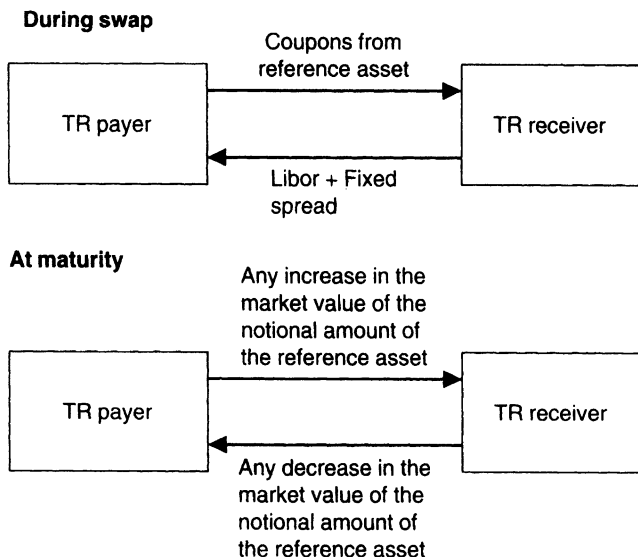


Fig. 3 Mechanics of a total return swap

Determination of the fixed spread to LIBOR payable by the total return receiver depends on several factors including the spread curve of the reference asset, the LIBOR curve, the financing cost to the payer of holding the asset on balance sheet, the expected default price of the asset, and counterparty credit quality. In certain circumstances, the price will be close to that of the corresponding default swap.

There are several reasons why an investor would wish to use such a structure. First and foremost is leverage. Using a total return swap, an investor can gain the return of an asset without paying the full price of the asset. The investor only has to make the coupon payments that are paid net. Indeed, if there is no actual transfer of the reference asset, then the limit on the size of the notional depends solely on the amount of risk that the two parties wish to assume.

Another motivation is that it enables investors to obtain off balance sheet exposure to assets that they might otherwise be precluded for tax, political or other reasons. This is especially useful to banks with lower credit ratings and higher funding costs in certain markets. Furthermore, total return swaps are often treated as derivatives and so incur a lower regulatory capital charge.

Total return swaps make it possible to short an asset without actually selling the asset. This may be useful from a point of view of temporarily hedging the risk of the credit, deferring a payment of capital gains tax, or simply gaining confidentiality regarding investment decisions.

As the maturity of the total return swap is not necessarily the same as the maturity of the reference asset, a total return swap effectively creates a new synthetic asset with the required maturity. Credit gaps in a portfolio may therefore be filled.

6.5 Principal Protection

Many loans to developing countries are like investing junk bonds, subject to high default risk. Banks that provide loans to developing countries are fully aware of the high default risk. Buying default swaps would be a natural way to hedge the risk but default swaps of these countries are usually very expensive and will wipe out all the incentive to provide loan to these countries. Furthermore, most banks that provide loan to these countries normally do so due to political and not economic considerations. Hence, protecting the principal and not the interest is the main concern for the lending banks.

A principal protected note (PPN) is developed for this purpose. PPN is similar to a risky floater except that the principal is guaranteed upon default and is similar to a risk-free floater except that the coupons are not guaranteed. As a result, we can value any PPN by either one of the following:

- Risky floater + Principal protection
- Risk-free floater – Risk-free coupons + Risky coupons
- Risk-free zero + Risky coupons

Since the default swaps hedge away coupon risks and the principal is risk-free, a PPN and the default swaps add up to a risk-free floater.

6.6 Credit Spread Options

A spread option is a contract whose payout depends on the credit spread of a reference asset. This reference asset may be either a floating rate note or an asset swap. As with standard options, one must specify whether the option is a call or put, the expiry date of the option, the strike price, and the type of optionality, i.e., American style (exercise at any time), European style (exercise at expiry only), or Bermudan style (exercise on one of several dates). It is also important to define what happens in the event that the underlying asset defaults – one may not want to pay for the right to exercise upon default. On exercise, the option may be settled through cash or physical delivery (Fig. 4).

Credit spread options are usually quoted in terms of the price of the reference asset. To see how an option on prices translates to an option on spreads, consider the case of a put option on a floating rate note paying LIBOR plus 150 basis points where the strike price is par. As the credit spread of the floating rate note widens, its price falls. If the market asset swap spread widens from 150 basis points to 160 basis points, the price of the floating rate bond falls below par and the put option is now in the money. The sensitivity of the price of the reference asset to changes in the asset swap spread is given by the spread duration. We can therefore view the put option on the price of the reference as a call option on the asset swap spread of the reference asset with an effective notional proportional to the spread duration.

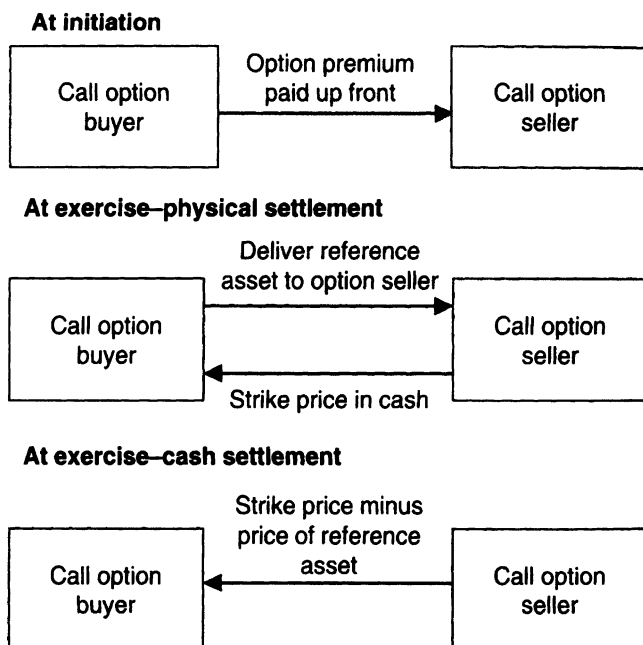


Fig. 4 Mechanics of a call option on reference asset

A more complicated version of this is the option on an asset swap shown in Fig. 5. In this case, the purchaser of a call option pays premium to the option seller to have the right to buy a specific reference asset and enter into a par flat asset swap.

As the asset swap spread of the reference asset widens, the value of the underlying asset swap falls and the buyer is less likely to exercise the call option. Equally, as the value of the underlying asset swap spread rises, the buyer is more likely to exercise the call option. Therefore, the call option on the asset swap translates into a put option on the asset swap spread and vice versa.

For all practical purposes, default swaps and credit spread options are the same. Indeed the payoff from a credit spread put option (a call on the asset swap spread) struck at par is the same as the payoff from a default swap, which pays par minus the default price. However, there are some differences. First of all, the value of a credit spread option depends on the credit spread volatility. The more volatile the credit spread, the more time-value the option will have and the more the option will be worth. Secondly, the payoff of credit spread option is sensitive to large increases in spread that may not actually constitute formal default. They therefore provide a hedge against price risk as well as default risk. Lastly, they allow the purchaser the right to choose when to exercise.

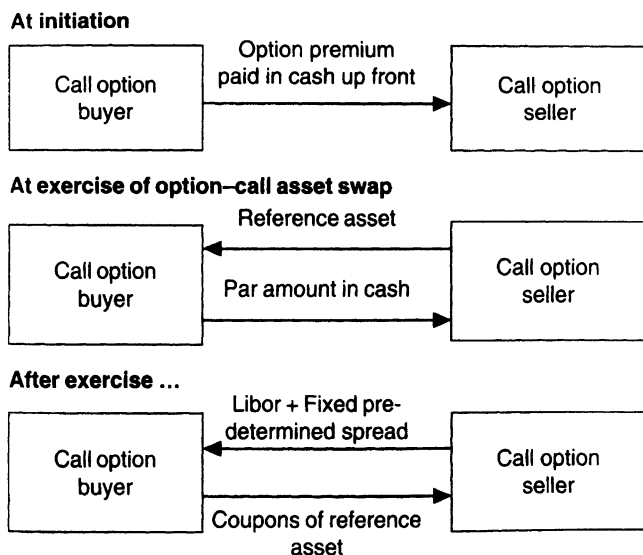


Fig. 5 Mechanics of a call option on asset swap

6.7 Basket Default Swaps

A basket default swap is like a default swap, but the only exception is that it is linked to the default of more than one credit. In the particular case of a first-to-default basket, it is the first asset in a basket whose credit event triggers a payment to the protection buyer. As in the case of a default swap, this payment may involve either cash payment of par minus the default price of the defaulted asset, or physical delivery of the defaulted asset in return for par. In return for protection against the first-to-default, the protection buyer pays a fee to the protection seller as a set of regular accruing cash flows.

To see clearly the mechanics of the structure, consider a deal in which an investor buys first-to-default protection to hedge a \$50M notional of each of three credits A, B, and C. Although the total notional amount covered is \$150M, it is imperative to note that if one of the credits is defaulted, only the notional size of that credit in the basket gets paid. For example, if credit B defaults, then we receive a payout equal in value to the difference between par and its default price on a notional of \$50M (see Fig. 6). The default basket terminates and remaining credits A and C are then left unhedged.

Since there is no simple portfolios that can statically replicate this structure, determining the value of the fixed spread is not an easy task. However, we can easily impose lower and upper limits on the price. Since the structure provides less protection than buying default swaps against *each* of the reference credits

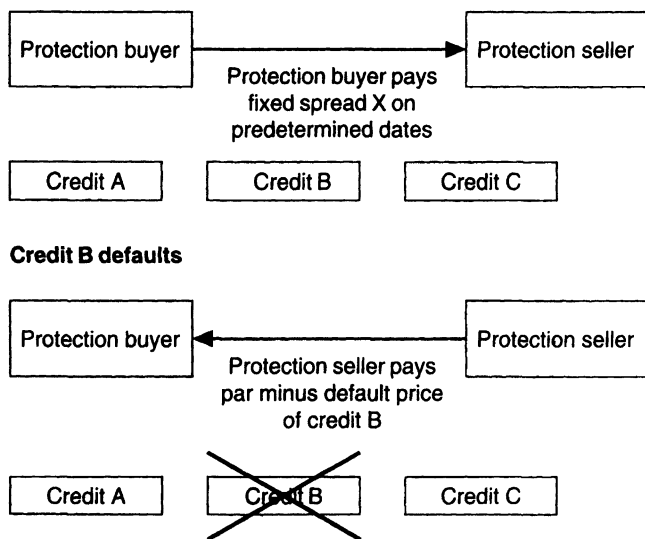


Fig. 6 A basket default swap on credits A, B and C

individually, it has to be less than this total cost. For a lower bound, we note that the cost of the basket has to be greater than the price of a default swap on the lowest credit quality asset in the basket. The problem is that these bounds may be quite far apart so that in practice we will need a model to get a more accurate price.

The motivation for doing a basket default swap is that it is a cheaper way to buy protection on a group of credits than buying protection individually on each credit. It is therefore an efficient way to reduce credit concentrations at an attractive cost. For the protection seller, the main motivation is that it provides a way to earn a high yield on high-quality securities.

There may also be a regulatory capital advantage to selling protection. For example, an investor may be able to sell protection against five assets, earning a high yield in the process, but only be required to pay the regulatory capital charge against one of the assets in the basket. However, as there is as yet no standard treatment for default baskets, the benefit of this advantage may vary depending upon regulatory framework.

It is important to understand that default baskets are correlation trades. However, there are two types of correlations to think about. First, there is the correlation between the changes in the spreads of the assets in the basket. This captures the fact that as one asset becomes more likely to default, another asset may also become more likely to default. The second type of correlation is the default correlation. This captures the knock-on effect that the default of one asset has on the default of another asset.

This is a subtle issue. To see it more clearly, consider the example of a default basket on two issuers within the same industry sector. We would expect to find a

strong positive correlation between the credit spread changes of both issuers. However, if one issuer were to actually default and this was due to idiosyncratic reasons, it has a beneficial effect on the other issuer due to effects such as creating more market share and reducing labor costs. The upshot is that we have positive spread correlation but negative default correlation. A major difficulty is the sheer lack of data available for estimating these correlations. In practice, the credits in most baskets are chosen in such a way that they have low probabilities of default and low correlations with each other.

6.8 Convertible Bonds

Convertible bonds are traditionally regarded as an equity play, i.e., the buyers of convertibles are after potentially high equity value. Some convertible bonds with less likelihood of significant equity appreciation try to attract buyers who try to enhance investment returns. Nowadays, convertibles have become a major credit derivative contract. Institutional investors discover that the credit risk in convertible bonds is significantly mispriced and hence start to arbitrage on high-yield convertibles.

A typical credit play of convertibles is to buy an underpriced convertible, engage in an interest rate swap to hedge away the interest rate risk, sell an equity put option to compensate the conversion value paid for, and enjoy the underpriced credit spread. This credit spread can be monetized by selling a spread option, or engaging in a forward asset swap, or buying a default swap.

6.9 Conclusions

What we have presented here constitute the main types of credit derivative instruments. Many of the more exotic structures that are now being traded in the market are simply variations and extensions of these basic building blocks. For example, credit-linked notes may consist simply of a standard bullet bond that has an embedded default swap – the investor receives a coupon plus a spread and loses part of the redemption value if the reference credit defaults.

One important issue in the use of credit derivatives, which is not the focus of this review, is how to price and hedge these instruments. See, for example, Caouette et al. (1998), Saunders and Allen (2002), Duffie and Singleton (2003), and Anson et al. (2004).

Acknowledgement We would like to thank Dominic O’Kane of Lehman Brothers for helpful comments and suggestions.

References

- Anson, M., F. Fabozzi, M. Choudhry, and R.R. Chen. 2004. *credit derivatives: instruments, applications, and pricing*. New York: Wiley.
- Caouette, J.B., E.I. Altman, and P. Narayanan. 1998. *Managing credit risk: The next great financial challenge*. New York: Wiley.
- Duffie, D., and K.J. Singleton. 2003. *Credit risk: Pricing, measurement and management*. Princeton: Princeton University Press.
- Saunders, A., and L. Allen. 2002. *Credit risk measurement*. 2nd ed. New York: Wiley.



Foreign Exchange Risk Premium and Policy Uncertainty

7

Thomas C. Chiang

Abstract

This chapter investigates the sources of bias by using a forward foreign exchange rate as a predictor of the corresponding spot exchange rate. The forward exchange rate bias is usually perceived to be associated with a risk factor. Numerous studies, which use variables such as a forward premium, conditional variance, real interest rate differential, and equity premium differential, have attempted to explain the forward-rate bias. This chapter conducts empirical tests that add a number of news-based risk variables in four foreign exchange markets to a proposed model. Testing results suggest that the news-based exchange rate volatility, equity premium differential, economic policy uncertainty differential, and in some cases, a geopolitical risk (GPR) are key factors to explain the forward-rate bias and therefore support the risk premium hypothesis.

Keywords

Forward exchange rate risk premium · Unbiased forward rate hypothesis · Equity premium · Real interest rate · Economic policy uncertainty · Geopolitical risk

JEL Code

F31 · F32

T. C. Chiang (✉)

Department of Finance, Drexel University, Philadelphia, PA, USA

e-mail: Chiangtc@Drexel.edu

7.1 Introduction

A simple way to predict the future spot exchange rate is to use forward exchange rate information (Levine 1989; Engel 1996, 2016). This approach is based on a rational expectation, which suggests that future information on exchange rate movements will be effectively reflected in the current forward exchange rate. For instance, the monetary approach to the exchange rate model posits that the exchange rate is determined by relative money supplies, relative incomes, and interest rate differentials, among others. Rational expectations imply that future information associated with those arguments will be reflected in the forward exchange rate, which is based on participants' effective assessments of the market. Should deviations exist, the difference between the future spot rate (s_{t+1}) and the current forward exchange rate (f_t), defined as the forward-rate bias, is usually attributable to a time-varying risk premium. A large number of studies have been dedicated to measuring of the risk factor. An early and highly popular approach, based on the measure of forward premium, has been tested by Hansen and Hodrick (1980), Hodrick (1987), Fama (1984), Chiang (1986, 1988), and Bekaert and Hodrick (1993), but the evidence for supporting this forward premium hypothesis is rather weak.

With the introduction of Engle's Autoregressive Conditional Heteroscedastic (ARCH) process of estimating exchange rate variance, Domowitz and Hakkio (1985), Baillie and Bollerslev (1990), and Chatterjee and Biswas (2009) find supporting evidence that an exchange rate risk premium is associated with exchange rate volatility. Evidence by Jiang and Chiang (2000) even indicates that the effect of stock market risk may spillover to the currency markets. Nevertheless, the conditional variance model often suffers from low performance when it is used to predict the risk premium because it fails to provide immediate information or is incapable of encompassing a sufficiently broad body of information. The availability of a news-based measure of forward exchange-rate market volatility (EMV) such as the index developed by Baker et al. (2019) fills the gap left by this missing information in a timely manner. But does this EMV measure achieve its goal? This chapter explores this question.

The search for some economic variables that could provide economic rationale to interpret the foreign exchange risk premium (FXRP) has occupied economists for some time. In his study of the pricing of forward contracts for foreign exchange, Korajczyk (1985) employs the real interest rate differential as the main argument to interpret the risk premium. In fact, the risk under this perspective essentially stems from the imminent uncertainty of inflation rate expectations relative to the interest rates. Obviously, this focus tends to ignore the activity occurring in stock markets. Follow-up studies by Chiang (1991) and Korajczyk and Viallet (1992) turn particular attention in this direction and propose that equity risk premiums are the key factor that explains the risk premium. From this perspective, risk is assumed to arise from stock market uncertainty. In bringing these two approaches together, it is important to include consideration for the uncertainty from international markets. Chiang (2006) and Chiang and Yang (2007), therefore, explore information in relation to the balance-of-payments transactions and find evidence that both the

equity premium differential and the real interest rate differential are significant in explaining risk premium.

Using above-mentioned literature as a backdrop for further investigation, this study attempts to explain FXRP by examining different model specifications for four different currency markets. Following this introduction, Sect. 7.2 discusses the data; Sect. 7.3 presents different econometric models by abstracting time series information; Sect. 7.4 presents an asset return differential model to predict FXRP; Sect. 7.5 introduces the news-based variables to the test equation and reports the related empirical evidence; Sect. 7.6 discusses the study's conclusions.

7.2 Data

The data in this study cover the currency prices for Canada (CA), European Union (EU), Japan (JP), and Switzerland (SW) against the US dollar for the period of June 1990 through December 2019. The selection of these markets is constrained by the availability of the economic policy uncertainty (EPU) (Baker et al. 2019) and geopolitical indices provided by Caldara and Iacoviello (2021). In empirical estimations, the US market is treated as a reference market due to its relatively dominant size and effectiveness in processing information. Thus, it allows us to examine the impact of US policy changes on each of the four currency markets. This study uses end-of-the-month spot exchange rates and one-month forward exchange rates, with the exchange rates expressed as prices of the local currency per unit of the US dollar. Short-term interest rates are measured by one-month euro-currency deposit rates for each country. These euro-currency deposit rates have been widely used in empirical studies due to their homogeneous features and the convenience in making comparisons across markets. The stock price indices are the “total stock market return index (RI in *Datastream*)” constructed by Data Stream. Stock returns are constructed by taking the log-difference. Inflation rates are measured by taking the log-difference of each market's CPI. All variables are measured on a monthly basis, which is in line with the fact that consumer price indices are available only on a monthly basis. By employing monthly observations, we are able to construct variables such as forward-exchange rates and short-term interest rates with the same maturity and avoid problems due to overlapping data.

The EPU is based on a three-word criterion established by Baker et al. (2016) for {E,P,U}: Economic, Policy, Uncertainty, and their variants, which are extracted from daily, local, and major local newspapers. The index is available and can be downloaded at the following link: https://www.policyuncertainty.com/global_monthly.html. Davis (2020) and his associates construct a scaled frequency count of articles related to EPU for each country. The Global Economic Policy Uncertainty (GEPU) Index is a balanced panel of monthly EPU index values for 21 countries from January 1997 to the present. The computed GEPU Index is applied to Switzerland in this study. Similarly, Baker et al. (2019) also construct a

measure of exchange rate market volatility (EMV) based on major newspapers around the world.

The Geopolitical Risk (GPR) Index stems from the work of Caldara and Iacoviello (2019). They define GPR as the “risk associated with wars, terrorist acts, and tensions between states that affect the normal and peaceful course of international relations.” They construct a monthly index by counting the occurrence of words related to geopolitical tensions in 11 leading international newspapers and search articles containing references to the following words or press coverage of actual adverse geopolitical events: {military-related tensions, nuclear tensions, war threats, terrorist threats, terrorist acts, or the beginning of a war}. Caldara and Iacoviello show that a GPR Index has little correlation with a EPU Index, except around 9/11 and the Iraq War. The GPR indices are available at <https://www.matteoiacoviello.com/gpr.htm>.

7.3 FXRP Models Based on Time Series Information

7.3.1 Forward Premium Hypothesis

The foreign exchange risk premium is usually defined as the difference between the expected future spot exchange rate (Hakkio and Sibert 1995), and forward exchange rate, expressed as $s_{t+1} - f_t$; these two variables are usually expressed as a natural logarithm. Bilson (1981) proposes a speculative efficient hypothesis, which argues that the future spot exchange rate can be forecasted by using current information of spot and forward exchange rates. It can be expressed as:

$$s_{t+1}^e = \alpha + w_1 f_t + w_2 s_t, \text{ where } w_1 + w_2 = 1 \quad (1)$$

where $w_1 + w_2 = 1$. Equation (1) states that the expected future spot exchange rate (s_{t+1}^e) can be predicted by a weighted average of forward (f_t) and current spot (s_t) exchange rates. By using rational expectations: $s_{t+1} = s_{t+1}^e + \varepsilon_{t+1}$, Eq. (1) can be expressed as:

$$s_{t+1} - f_t = \alpha + \beta_1 (f_t - s_t) + \varepsilon_{t+1}, \quad (2)$$

$$\text{or } s_{t+1} - s_t = \alpha + \beta_1' (f_t - s_t) + \varepsilon_{t+1}, \quad (2')$$

where $\beta_1 = -(1 - w_1)$, $\beta_1' = (1 + \beta_1)$. Equation (2) states that the forward premium, $f_t - s_t$, can be used to predict the forward exchange rate risk and, Eq. (2') can be used to predict the future spot rate change. Panel A in Table 1 reports the regression results using Newey-West (1987) HAC (heteroskedasticity and autocorrelation consistent) estimator. Evidence suggests that three out of four markets have the expected signs where β_1 is negative; however, none of these coefficients is statistically significant. The statistics in Panel B show mixed results. However, the evidence did turn out to be consistent with the restriction that $\beta_1' = (1 + \beta_1)$.

Table 1 Estimates of the forward premium as a predictor for FX risk premium

$A. s_{t+1} - f_t = \alpha + \beta_1(f_t - s_t) + \varepsilon_{t+1}$				
	α	β_1	DW	R^2
CA	-0.0001	-2.1249	2.0838	0.0044
	-0.08	-1.09		
EU	0.0001	-0.3858	1.9161	0.0010
	0.03	-0.46		
JP	0.0007	0.0091	1.9231	0.0000
	0.38	0.02		
SW	0.0000	-0.0152	2.1228	0.0000
	0.02	-0.03		
$B. s_{t+1} - s_t = \alpha + \beta'_1(f_t - s_t) + \varepsilon_{t+1}$				
	α	β'_1	DW	R^2
CA	-0.0001	-1.1249	2.0838	0.0012
	-0.08	-0.57		
EU	0.0001	0.6142	1.9161	0.0026
	0.03	0.74		
JP	0.0007	1.0091	1.9231	0.0137
	0.38	2.47		
SW	0.0000	0.9848	2.1228	0.0113
	0.02	1.99		

This table presents evidence that uses a forward premium ($f_t - s_t$) to predict risk forward exchange rate risk premium. The estimated equation in Panel A is: $s_{t+1} - f_t = \alpha + \beta_1(f_t - s_t) + \varepsilon_{t+1}$ and variance equation is $\sigma_t^2 h_{i,t} = \omega_0 + \omega_1 \varepsilon_{t-1}^2 + \omega_2 \sigma_{t-1}^2$. Panel B estimates $s_{t+1} - s_t = \alpha + \beta'_1(f_t - s_t) + \varepsilon_{t+1}$. For each market, the first row reports the estimated coefficients, the second row contains the estimated t-statistics. The critical values of t-distribution at the 1%, 5%, and 10% levels of significance are 2.63, 1.98, and 1.66, respectively. R^2 is the R-squared for the coefficient of determination. $\beta'_1 = 1 + \beta_1$. 0.0000 indicates a very small number.

A review of the R-squares shows that the values are rather low, indicating poor performance of the forward premium model. The estimated results are consistent with earlier findings reported by Hansen and Hodrick (1980), Hodrick (1987), Fama (1984), and Chiang (1986, 2006).

7.3.2 Conditional Variance

A shortcoming of the forward rate hypothesis is its inability to explicitly include a risk factor for predicting the FXRP. Spurred by Engle's ARCH model specification, economists assume that the missing variable is a measure for exchange rate volatility. As modeled by Domowitz and Hakkio (1985), and Jiang and Chiang (2000), this variable is associated with the conditional variance as projected by its lagged variance and past shock squares. The system can be specifically expressed as:

Table 2 Estimates of conditional variance as a predictor for FX risk premium

	α	β_2	$MA(q)$	ω_0	ω_1	ω_2	R^2
CA	0.0082	−6.2363	−0.072(1)	0.0004	0.1191	0.6383	0.0017
	0.80	−0.67	−3.00	1.11	0.67	2.37	
EU	0.0150	−15.2491	0.123(6)	0.0000	0.0562	0.9144	0.0147
	2.01	−1.71	2.54	1.15	1.86	22.71	
JP	0.0038	−2.1223	−0.118(5)	0.0004	0.0949	0.5376	0.0217
	0.57	−0.33	−3.51	0.58	0.64	0.74	
SW	−0.0204	6.1195	—	0.0008	0.1736	0.7465	0.0154
	−4.21	2.60	—	3.32	2.28	63.17	

This table presents evidence that uses the conditional volatility of exchange rate, $(\hat{\sigma}_{s,t+1}^2)$, by a GARCH(1,1) model to predict risk premium of exchange rate, $(s_{t+1} - f_t)$. The estimated equation is: $s_{t+1} - f_t = \alpha + \beta_2 (\hat{\sigma}_{s,t+1}^2) + \varepsilon_{t+1}$ and variance equation is $\sigma_t^2 = \omega_0 + \omega_1 \varepsilon_{t-1}^2 + \omega_2 \sigma_{t-1}^2$. MA (q) is the Moving Average terms, the numbers in the parentheses indicate the order of lagged terms for MA; evidence shows MA(1) for CA, MA(6) for EU, and MA(5) for JP and absence of MA for SW. For each market, the first row reports the estimated coefficients and the second row contains the estimated t-statistics. The critical values of t-distribution at the 1%, 5%, and 10% levels of significance are 2.63, 1.98, and 1.66, respectively. R^2 is the R-squared for the coefficient of determination. 0.0000 indicates a very small number

$$s_{t+1} - f_t = \alpha + \beta_2 (\hat{\sigma}_{s,t+1}^2) + \varepsilon_{t+1} \tag{3}$$

$$\text{with } \sigma_{t+1}^2 = \omega_0 + \omega_1 \varepsilon_t^2 + \omega_2 \sigma_t^2. \tag{3'}$$

The system of Eqs. (3) and (3') posits that the FXRP is associated with the conditional variance developed by a GARCH(1,1) process.¹ The null is $\beta_2 = 0$. Eq. (3') states that the variance evolves with past shock squares and previous variance. The estimated results, which use GARCH(1,1) procedure while controlling MA(q) process in the mean equation, are reported in Table 2. The estimated coefficients of β_2 display mixed signs with Switzerland exhibiting a positive sign and the other three markets negative signs. The low performance of the conditional variance model does not match the notion of mean-variance prediction. This could result from inappropriate frequency of the data or could be attributable to the failure of historical information for projecting risk information.

7.3.3 News-Based Exchange Rate Volatility Hypothesis

One possible explanation for low performance of the conditional variance variable is that it relies on a single source of information and fails to capture on-going news developments. As noted by Shiller: “The history of thought in financial markets has shown a surprising lack of consensus about a very fundamental question: what

¹This model specification was popularized by Bollerslev et al. (1992). Different model specifications for ARCH models can be found in Bollerslev (2010).

Table 3 Estimates of change for exchange rate market volatility as predictor for FX risk premium

	α	β_3	ω_0	ω_1	ω_2	R^2
CA	0.0000	0.0005	0.0001	0.0655	0.9494	0.0044
	-0.10	5.99	0.68	0.36	11.46	
EU	0.0002	0.0006	0.0015	1.9334	0.8501	0.0058
	0.94	6.03	0.30	0.43	2.50	
JP	0.0007	0.0006	0.0002	0.2759	0.8570	0.0039
	2.63	4.58	0.43	0.63	3.62	
SW	0.0009	0.0004	0.0002	0.3683	0.8421	0.0010
	3.35	2.92	0.51	0.66	3.72	

This table presents evidence that uses a change in news-based volatility of exchange rate market, ΔEMV_t , to predict risk premium of exchange rate ($s_{t+1} - f_t$). The estimated equation is: $s_{t+1} - f_t = \alpha + \beta_3(\Delta EMV_t) + \varepsilon_{t+1}$ and variance equation is $\sigma_t^2 h_{i,t} = \omega_0 + \omega_1 \varepsilon_{t-1}^2 + \omega_2 \sigma_{t-1}^2$. For each market, the first row reports the estimated coefficients, and the second row contains the estimated t-statistics. The critical values of t-distribution at the 1%, 5%, and 10% levels of significance are 2.63, 1.98, and 1.66, respectively. R^2 is the R-squared for the coefficient of determination

ultimately causes all those fluctuations in the price of speculative assets like corporate stocks. ...?” (Shiller 2014). One way to improve the information content is to employ a newspaper-based measure of exchange rate volatility (EMV_t) developed by Baker et al. (2019). Unlike the traditional news-approach model based on an unexpected component of a state variable (Edwards 1982; Chiang 1985), this index covers a broad body of information pertaining to exchange rates and their variants, such as {exchange rate}, {currency crisis}, {currency devaluation, currency depreciation}, {currency revaluation, currency appreciation}, {crawling peg, managed float}, {currency manipulation, currency intervention} cited in major newspapers. This leads to a model specified as:

$$s_{t+1} - f_t = \alpha + \beta_3(\Delta EMV_t) + \varepsilon_{t+1} \tag{4}$$

where ΔEMV_t is a broad index variable that measures a change in exchange market volatility. Table 3 presents evidence that uses a change in exchange rate market volatility to predict FXRP when the GED-GARCH(1,1) procedure (Nelson 1991) is employed. The coefficient β_3 is positive and statistically significant, which is consistent with the notion that a rise in FXRP correlates to an increase in exchange rate volatility. Evidence allows us to conclude that the use of ΔEMV_t as a risk argument in Table 3 appears to outperform the use of the conditional variance because the former conveys exchange rate volatility information in a more instant, efficient, and broader manner than the conditional volatility.

7.4 Predicting FXRP Based on Asset Return Differentials

7.4.1 Equity Risk Premium Differential Hypothesis

To explore the uncertainty from financial market, Chiang (1991) and Korajczyk and Viallet (1992) use the relative equity market premium as a proxy for two country’s risk differentials. The expression is given by:

$$s_{t+1} - f_t = \alpha + \gamma_1 [(R_{t+1} - r_t) - (R_{t+1}^* - r_t^*)] + \varepsilon_{t+1} \tag{5}$$

where $R_{t+1} - r_t$ and $R_{t+1}^* - r_t^*$ are, respectively, equity risk premium in domestic and foreign countries.² The γ_1 is a constant coefficient. It readily shows that:

$$s_{t+1} - f_t = \alpha + \gamma_1 [(R_{t+1} - R_{t+1}^*) - (r_t - r_t^*)] + \varepsilon_{t+1} \tag{5'}$$

A special feature of this model is its focus on the risk as reflected in the stock markets and indicated by $(R_{t+1} - R_{t+1}^*)$. In this expression, the bond market returns $(r_t - r_t^*)$ are considered to contain less risk. The estimated results of Eq. (5) are reported in Table 4, which indicates that the coefficients of γ_1 are consistently positive and statistically significant. The results are consistent with the anticipated theory that a country's relatively high risk leads to a depreciation in its national currency. For this reason, the use of a current forward rate as a predictor will create a biased estimator for the future spot rate. The results are consistent with the findings by Chiang (1991, 2006).

7.4.2 Real Interest Rate Differential Hypothesis

Influenced by the study of Frankel (1979), who develops a theory of predicting floating exchange rates that are based on real interest differential, Korajczyk (1982) utilizes the real interest rate differential to explain the FXRP. The model can be expressed as:

$$s_{t+1} - f_t = \alpha + \gamma_2 [(r_t - \Delta p_{t+1}^e) - (r_t^* - \Delta p_{t+1}^{*e})] + \varepsilon_{t+1} \tag{6}$$

Table 4 Estimates of the equity risk premium differential as predictor for FX risk premium

	α	γ_1	AR(4)	AR(5)	ω_0	ω_1	ω_2	R^2
CA	0.0007	0.2432			0.0000	0.0384	0.9329	0.0138
	0.54	2.19			1.20	1.83	24.72	
EU	0.0003	0.0196	0.2699		0.0000	0.0465	0.8874	0.0605
	1.48	2.05	4.00		1.21	1.32	11.90	
JP	0.0018	0.4333		-0.1346	0.0000	0.0443	0.9489	0.0675
	1.44	7.19		-2.50	0.41	2.27	44.05	
SW	-0.0004	0.8719			0.0000	0.0717	0.9009	0.1583
	-0.32	9.52			0.87	1.82	15.71	

The estimated equation is $s_{t+1} - f_t = \alpha + \gamma_1 [(R_{t+1} - r_t) - (R_{t+1}^* - r_t^*)] + \varepsilon_{t+1}$. For each market, the first row reports the estimated coefficients and the second row contains the estimated t-statistics. The critical values of t-distribution at the 1%, 5%, and 10% levels of significance are 2.63, 1.98, and 1.66, respectively. R^2 is the R-squared for the coefficient of determination. 0.0000 indicates a very small number

² An asterisk over a variable denotes foreign variable, which denotes the U.S. in our setting.

where $(r_t - \Delta p_{t+1}^e)$ and $(r_t^* - \Delta p_{t+1}^{e*})$ are, respectively, real interest rates in a domestic and foreign countries. Rewriting this equation, it can be shown that:

$$s_{t+1} - f_t = \alpha + \gamma_2 [(r_t - r_t^*) - (\Delta p_{t+1}^e - \Delta p_{t+1}^{e*})] + \varepsilon_{t+1} \quad (6')$$

where $\gamma_2 < 0$. By imposing the covered interest rate parity condition, $f_t - s_t = r_t - r_t^*$, Eq. (2) can be written as:

$$s_{t+1} - f_t = \alpha + \beta_1 (r_t - r_t^*) + \varepsilon_{t+1} \quad (2^\dagger)$$

The resulting Eq. (2[†]) implies that the forward premium hypothesis is essentially linked to the FXRP via the interest rate differentials in the bond market. A comparison with Eq. (6') shows that the information for the inflation rate differential via trade account is missing, while a comparison of (2[†]) with Eq. (5') shows the information for the stock return differential is ignored in the forward premium hypothesis. Thus, either the real interest rate differential model in Eq. (6) or the equity premium differential model in Eq. (5) is richer because these models have more information content.

Estimates of Eq. (6), which applies the GED-GARCH(1,1) procedure (Nelson 1991), are reported in Table 5. Despite Korajczyk's (1982) successful demonstration of the original model, the updated data in Table 6 fail to develop evidence that fully supports the real interest rate hypothesis. This is reflected in the inconsistency in results that show three out of four markets bear a negative sign and are statistically significant, while the sign for Japanese market does not show a consistent one. This could result from a rigid interest rate policy in the Japanese market over the past two decades and, more generally, may hold true for other markets, that is, low performance of the real interest rate hypothesis could result from a relative rigidity of interest rates and inflation rate expectations when compared with the exchange rate premium.

7.4.3 Equity Premium Differential Cum Real Interest Rate Differential Hypothesis

A careful review of the underlying arguments using the above two models shows that either model is obviously subjected to a partial equilibrium approach. Based on the Balance-of-Payment theory, it can be readily argued that the expected change in exchange rate is mainly governed by the elements underlying the goods account and capita account as dictated by the expected inflation rate differentials, interest rate differential, and expected stock return differential. Thus,

$$\Delta s_{t+1}^e = w_1 (p_{t+1}^e - p_{t+1}^{e*}) + w_2 (r_t - r_t^*) + w_3 (R_{m,t+1}^e - R_{m,t+1}^{e*}) \quad (7)$$

where $w_1 + w_2 + w_3 = 1$ due to balance of payments constraint. By subtracting $r_t = r_t^*$ from both sides of the equation, based on the covered interest rate parity condition, $f_t - s_t = r_t - r_t^*$, and rearranging the terms yields:

Table 5 Estimates of the real interest rate differential as predictor of FX risk premium

	α	γ_2	AR(4)	AR(5)	ω_0	ω_1	ω_2	R^2
CA	0.0009	-0.8954			0.0000	0.0516	0.9398	-0.0121
	0.63	-2.16			1.39	4.23	55.61	
EU	0.0004	-0.0333	0.0602		0.0000	0.5523	0.8698	0.0322
	11.55	-3.16	5.09		0.27	0.58	3.36	
JP	0.0013	0.4621		-0.0078	0.0007	0.2811	0.6755	0.0031
	4.30	4.24		-2.88	0.47	0.49	1.11	
SW	-0.0009	-0.9661			0.0001	0.1323	0.8155	0.0040
	-0.87	-3.36			0.81	0.98	4.72	

The estimated equation is $s_{t+1} - f_t = \alpha + \gamma_2 [(r_t - \Delta p_{t+1}^e) - (r_t^* - \Delta p_{t+1}^{*e})] + \varepsilon_{t+1}$. For each market, the first row reports the estimated coefficients, the second row contains the estimated t-statistics. The critical values of t-distribution at the 1%, 5%, and 10% levels of significance are 2.63, 1.98, and 1.66, respectively. R^2 is the R-squared for the coefficient of de nation. 0.0000 indicates a very small number.

Table 6 Estimates of equity risk premium differential and real interest rate differential as predictors

	α	γ_1	γ_2	ω_0	ω_1	ω_2	R^2
CA	0.0004	0.3527	-0.3381	0.0002	0.2434	0.8760	0.0107
	1.15	10.54	-3.97	0.49	0.49	4.05	
EU	0.0002	0.0223	-0.0174	0.0000	0.0528	0.9143	0.0139
	1.71	2.32	-0.64	0.94	1.16	14.66	
JP	0.0013	0.3449	0.3640	0.0000	0.3324	0.9215	0.0344
	5.63	19.46	5.20	-0.10	1.07	11.03	
SW	0.0001	0.9061	-0.4192	0.0000	0.0896	0.9115	0.1500
	-0.01	12.38	-4.28	0.47	0.91	8.47	

This table presents evidence that uses the equity risk premium differential $[(R_{t+1} - r_t) - (R_{t+1}^* - r_t^*)]$ and real interest rate differential $[(r_t - \Delta p_{t+1}^e) - (r_t^* - \Delta p_{t+1}^{*e})]$ to predict the risk premium of exchange rate $(s_{t+1} - f_t)$. The estimated equation is $s_{t+1} - f_t = \alpha + \gamma_1 [(R_{t+1} - r_t) - (R_{t+1}^* - r_t^*)] + \gamma_2 [(r_t - \Delta p_{t+1}^e) - (r_t^* - \Delta p_{t+1}^{*e})] + \varepsilon_{t+1}$. For each market, the first row reports the estimated coefficients and the second row contains the estimated t-statistics. The critical values of t-distribution at the 1%, 5%, and 10% levels of significance are 2.63, 1.98, and 1.66, respectively. R^2 is the R-squared for the coefficient of determination. 0.0000 indicates a very small number.

$$s_{t+1}^e - f_t = w_3 \left[(R_{m,t+1}^e - r_t) - (R_{m,t+1}^{*e} - r_t^*) \right] + w_1 [(p_{t+1}^e - r_t) - (p_{t+1}^{*e} - r_t^*)]$$

Applying rational expectations, $s_{t+1} = s_{t+1}^e + \varepsilon_{t+1}$, and arranging the expression in a regression form yields:

$$s_{t+1} - f_t = \alpha + \gamma_1 \left[(R_{m,t+1}^e - r_t) - (R_{m,t+1}^{*e} - r_t^*) \right] + \gamma_2 [(r_t - \Delta p_{t+1}^e) - (r_t^* - \Delta p_{t+1}^{*e})] + \varepsilon_{t+1}, \quad (8)$$

where $\gamma_1 = \omega_3$, $\gamma_2 = -\omega_1$. By adding a constant term, we arrive Eq. (8).

Several points related to Eq. (8) are worth noting. First, the FXRP can be explained by both an equity premium differential and a real interest rate differential. Second, the sources of risk are associated with the uncertainty of stock returns and inflation rates in the future. Third, this model is a general representation of either Chiang's model (1991) or Korajczyk's model (1985) since Eq. (5) or Eq. (6) is nested in Eq. (8).

The estimated results of Eq. (8) are reported in Table 6. Evidence shows that the coefficients for the equity premium differentials remain positive and are statistically significant. These results are comparable to the earlier estimates reported in Table 4. However, the coefficients for the real interest rate differential now present mixed signs, that is, the coefficients show a negative sign and are significant for Canada and Switzerland (EU is insignificant); however, it is positive for the Japanese market. Testing results suggest that the equity risk differential argument has a more consistent performance.

7.5 Effects of Policy Uncertainty

7.5.1 Economic Policy Uncertainty

Although the above analyses cover sufficient information for modeling the FXRP, it fails to update the literature by including a measure for news-based volatility and economic policy uncertainty. These two indices are constructed by Baker et al. (2019) and Baker et al. (2016), respectively, and are updated from time to time. The link is available at following website https://www.policyuncertainty.com/us_monthly.html

$$s_{t+1} - f_t = \alpha + \gamma_1 [(R_{t+1} - r_t) - (R_{t+1}^* - r_t^*)] + \gamma_2 [(r_t - \Delta p_{t+1}^e) - (r_t^* - \Delta p_{t+1}^{*e})] + \beta_3 (\Delta EMV_t) + \delta_1 (EPU_t - EPU_t^*) + \varepsilon_{t+1} \quad (9)$$

In addition to the economic fundamentals of FXRP mentioned earlier, a special feature of this model is the inclusion of an equity risk premium differential, a real interest rate differential, and exchange market volatility (ΔEMV_t). Equation (9) also adds news-based relative economic policy uncertainty (Bloom, 2009, 2014), which is the difference between local and US EPU ($EPU_t - EPU_t^*$), where EMV_t , EPU_t , and EPU_t^* are expressed in natural logarithms. In priori, $\delta_1 > 0$ indicates that a country's uncertainty is higher as compared to that for the USA and is anticipated to depreciate. This uncertainty component is distinct from the risk factor of exchange rate volatility (ΔEMV_t).

The estimated results, which are reported in Table 7, are consistent with the findings we presented in earlier tables. The coefficients of exchange rate risk, equity

premium differential and real interest rate differential have the signs, are statistically significant and are comparable to those presented in Tables 3, 4 and 5. All estimated coefficients for $(EPU_t - EPU_t^*)$ are positive and statistically significant. The higher values of adjusted R-squares provide further evidence that supports the incremental efficiency, which results by adding the relative EPU as an explanatory variable. This element, however, has not been commonly incorporated into models in predicting FXRP.

7.5.2 Effect of Geopolitical Risk

Whether the forward rate can precisely predict future spot rate to some extent depends very much on the stability of the currency market, which is further contingent on the financial system stability and political environment. As noted by Trask (2017): “if a country’s economy is performing well and the future looks prosperous and - crucially - predictable, investors will be happy to hold its currency. Unexpected events, however, will send them running for the exit.” The reality is that traders’ behavior or investors’ sentiment has been subjected to the social stability. Currency values can be affected by a wide range of political events including general elections (such as vote for Brexit in UK), unexpected policy announcements (Trump’s trade policy with China), military threat (to Iran), or terrorist attacks (911). Clare and Gang find that decreasing political risk brings about a higher foreign direct investment (2010). Using data from eight industrial democracies between 1974 and 1995, Bernhard and Leblang (2002) document that a foreign exchange risk premium exists more often during political processes. For this reason, this study further examines the impact of GPR on the FXRP by adding a GPR variable to the test equation. Equation (9) becomes:

$$s_{t+1} - f_t = \alpha + \gamma_1 [(R_{t+1} - r_t) - (R_{t+1}^* - r_t^*)] + \gamma_2 [(r_t - \Delta p_{t+1}^e) - (r_t^* - \Delta p_{t+1}^{*e})] + \beta_3 (\Delta EMV_t) + \delta_1 (EPU_t - EPU_t^*) + \delta_2 (GPR_t) + \varepsilon_{t+1} \quad (10)$$

where GPR_t is the geopolitical risk in natural logarithm, other variables are defined previously. The estimated results of Eq. (10) are reported in Table 8.

Testing results suggest that the coefficients for the equity premium differential, real interest rate differential, the change of exchange rate market volatility, and the economic policy differential produce very comparable results as those presented in Table 7. However, the null of FXRP is independent of GPR is rejected for three out of four markets. The exception is Switzerland. In particular, Canada and the EU display a negative sign. The results are consistent with the market behavior that shows as the US GPR rises, anticipation regarding the spread in uncertainty would increase, disturbing financial markets, which in turn impedes economic activity that has a negative effect on the US dollar and leads to a depreciation of the US dollar. Using the forward rate to predict future spot rate would result in a bias under this

Table 7 Estimates of equity risk premium differential, real interest rate differential, exchange rate volatility and economic policy differential as predictors

	α	γ_1	γ_2	β_3	δ_1	AR(2)	AR(4)	ω_0	ω_1	ω_2	$\bar{R}^2$
CA	-0.0014 -9.17	0.3914 15.45	-0.0783 -3.28	0.0004 4.04	0.0068 8.55			0.0001 0.47	0.1060 0.46	0.8773 3.65	0.0194
EU	0.0000 -0.11	0.0096 2.02	-0.0043 -0.53	0.0001 4.91	0.0009 3.33	0.2526 9.16	0.2897 8.38	0.0000 0.68	0.0718 0.79	0.8422 4.47	0.0787
JP	0.0014 3.07	0.3022 9.65	0.3809 12.41	0.0006 4.61	0.0021 6.30			0.0000 0.12	0.3001 0.90	0.9202 9.16	0.0387
SW	-0.0008 -2.40	0.8619 34.76	-0.3296 -8.73	0.0002 17.40	0.0020 3.07			0.0005 0.38	0.2365 0.25	0.9226 5.59	0.1514

This table presents evidence that uses the equity risk premium differential $[(R_{t+1} - r_t) - (R_{t+1}^* - r_t^*)]$, real interest rate differential $[(r_t - \Delta r_{t+1}^e) - (r_t^* - \Delta r_{t+1}^{*e})]$, changes in domestic and foreign economic policy uncertainties, ΔEPU_t and ΔEPU_t^* , and change in news-based volatility of exchange rate, ΔEMV_t to predict the risk premium of exchange rate $(s_{t+1} - f_t = \alpha + \gamma_1[R_{t+1} - r_t] - (R_{t+1}^* - r_t^*)) + \gamma_2[(r_t - \Delta r_{t+1}^e) - (r_t^* - \Delta r_{t+1}^{*e})] + \beta_3(\Delta EMV_t) + \delta_1(EPU_t - EPU_t^*) + \varepsilon_{t+1}$. For each market, the first row reports the estimated coefficients, the second row contains the estimated t-statistics. The critical values of t-distribution at the 1%, 5%, and 10% levels of significance are 2.63, 1.98, and 1.66, respectively. R^2 is the R-squared for the coefficient of determination. 0.0000 indicates a very small number.

Table 8 Estimates of equity risk premium differential, real interest rate differential, exchange rate volatility, relative EPU differential, and geopolitical risk as predictors

	α	γ_1	γ_2	β_3	δ_1	δ_2	ω_0	ω_1	ω_2	$\bar{R}^2$
CA	0.0177	0.4884	-0.1995	0.0004	0.0055	-0.0041	0.0000	0.0564	0.9323	0.0110
	4.16	8.40	-3.56	2.78	5.32	-4.17	0.81	0.78	13.01	
EU	0.0026	0.0213	-0.0341	0.0000	0.0003	-0.0005	0.0000	0.5603	0.8418	0.0187
	9.23	6.84	-7.49	5.78	10.32	-9.36	0.29	0.68	3.53	
JP	-0.0051	0.3058	0.3511	0.0006	0.0006	0.0015	0.0005	0.4923	0.7840	0.0359
	-2.94	17.67	9.50	4.93	5.19	3.43	0.48	0.58	2.30	
SW	-0.0047	0.8973	-0.5493	0.0007	0.0012	0.0008	0.0005	0.3712	0.7541	0.1433
	-1.81	40.87	-4.65	11.21	16.51	1.39	0.41	0.49	1.54	

This table presents evidence that uses the equity risk premium differential $[(R_{t+1} - r_t) - (R_{t+1}^* - r_t^*)]$ real interest rate differential $[(r_t - \Delta p_{t+1}^e) - (r_t^* - \Delta p_{t+1}^{*e})]$, changes in domestic and foreign economic policy uncertainties, ΔEPU_t and ΔEPU_t^* , change in news-based volatility of exchange rate, ΔEMV_t , and geopolitical risk, GPR_t , to predict the risk premium of exchange rate $(s_{t+1} - f_t)$. The estimated equation is $s_{t+1} - f_t = \alpha + \gamma_1 [(R_{t+1} - r_t) - (R_{t+1}^* - r_t^*)] + \gamma_2 [(r_t - \Delta p_{t+1}^e) - (r_t^* - \Delta p_{t+1}^{*e})] + \beta_3 (\Delta EMV_t) + \delta_1 (EPU_t - EPU_t^*) + \delta_2 (GPR_t) + \varepsilon_{t+1}$. For each market, the first row reports the estimated coefficients and the second row contains the estimated t-statistics. The critical values of t-distribution at the 1%, 5%, and 10% levels of significance are 2.63, 1.98, and 1.66, respectively. $\bar{R}^2$ is the R-squared for the coefficient of determination. 0.0000 indicates a very small number.

scenario. A bias is also present in the Japanese market; however, the bias moves in the opposite direction. Factors that might produce mixed signs may be clouded by including only one measure of GPR. In fact, since exchange rate determination involves two national currencies, the argument that GPR should be modeled in a relative fashion is an effort, which would be better covered in future research.

7.6 Conclusions

By now a substantial body of evidence has suggested that the unbiased forward rate hypothesis is rejected. The divergence between the expected future spot rate and current forward rate is considered as a forward exchange rate risk premium, which is related to a measure of risk. This study examines several measures of risk, including the forward premium, conditional variance of exchange risk, exchange rate volatility, real interest rate differential, equity premium differential, economic policy uncertainty differential, and geopolitical risk. Testing of four different currency markets, including Canada, European Union, Japan, and Switzerland, produces consistent empirical results which are summarized as follows.

First, in a simple regression estimation, the data do not support the forward premium hypothesis, nor can we find strong evidence for the conditional variance model. Second, the statistics consistently support the use of news-based exchange rate volatility as a measure for predictability, although its explanatory power is not very high. Third, the evidence cannot find uniform support for using the real interest rate differential to predict a forward exchange risk premium. However, the equity premium differential provides good supporting evidence to predict the forward

exchange risk premium. Fourth, in a multiple regression setting, evidence favors the use of the equity premium differential, relative economic policy uncertainty, and news-based exchange rate volatility in predicting a risk premium, since the evidence provides consistent and significant results. In some cases, the geopolitical risk could also contribute to explaining the forward exchange risk premium.

References

- Baillie, R.T., and T. Bollerslev. 1990. A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets. *Journal of International Money and Finance* 9: 309–324.
- Baker, S.R., N. Bloom, and S. Davis. 2016. Measuring economic policy uncertainty. *Quarterly Journal of Economics* 131: 1593–1636.
- Baker, S., Bloom, N., Davis, S.J., and Kost, K. (2019). Policy news and stock market volatility. These data can be obtained from www.PolicyUncertainty.com.
- Bekaert, G., and R.J. Hodrick. 1993. On biases in the measurement of foreign exchange risk premiums. *Journal of International Money and Finance* 12: 115–138.
- Bernhard, W., and D. Leblang. 2002. Democratic processes, political risk, and foreign exchange markets. *American Journal of Political Science* 46 (2002): 316–333.
- Bilson, J.F.O. 1981. The ‘speculative efficiency’ hypothesis. *Journal of Business* 54: 435–451.
- Bloom, N. 2009. The impact of uncertainty shocks. *Econometrica* 77: 623–685.
- . 2014. Fluctuations in uncertainty. *Journal of Economic Perspectives* 28: 153–176.
- Bollerslev, T. 2010. Glossary to ARCH (GARCH). In *Volatility and time series econometrics: Essays in honor of Robert Engle*, ed. Tim Bollerslev, Jeffrey Russell, and Mark Watson. Oxford: Oxford University Press.
- Bollerslev, T., R.Y. Chou, and K.F. Kroner. 1992. ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52: 5–59.
- Caldara, D., Iacoviello, M., 2021. Measuring Geopolitical Risk. Working paper, Board of Governors of the Federal Reserve. <https://www.matteoiacoviello.com/gpr.htm>
- Chatterjee, D., and B. Biswas. 2009. Explaining forward rate unbiasedness hypothesis: The risk premium approach. *Metamorphosis – A Journal of Management Research* 8 (2): 19–30. <https://doi.org/10.1177/0972622520090204>.
- Chiang, T.C. 1985. The Impact of unexpected macro-disturbances on exchange rates in monetary models. *The Quarterly Review of Economics and Business* 25: 49–60.
- . 1986. On the predictors of the future spot rate – A multi-currency analysis. *Financial Review* 21: 69–84.
- . 1988. The forward rate as a predictor of the future spot rate – A stochastic coefficient approach. *Journal of Money, Credit and Banking* 20: 210–232.
- . 1991. International asset pricing and equity market risk. *Journal of International Money and Finance* 10: 349–364.
- . 2006. International parity conditions and market risk. In *Encyclopedia of finance*, ed. C.F. Lee and A.C. Lee, 344–358. New York: Springer Publishers.
- Chiang, T.C., and S.-Y. Yang. 2007. Foreign exchange risk premiums and relative asset return differentials. *International Research Journal of Finance and Economics* 8: 181–195.
- Clare, G., and I. Gang. 2010. Exchange rate and political risks, again. *Emerging Markets Finance & Trade* 46: 46–58.
- Davis, S. J., 2020. An index of global economic policy uncertainty. *NBER Working Paper* No. 22740. Updated of Davis, 2016. <https://www.nber.org/papers/w22740>
- Domowitz, I., and C.S. Hakkio. 1985. Conditional variance and the risk premium in the foreign exchange market. *Journal of International Economics* 19: 47–66.

- Edwards, S. 1982. Exchange rates and 'news': A multi-currency approach. *Journal of International Money and Finance* 1: 211–224.
- Engel, C. 1996. The forward discount anomaly and the risk premium: A survey of recent evidence. *Journal of Empirical Finance* 3: 123–192.
- . 2016. Exchange rates, interest rates, and the risk premium. *American Economic Review* 106: 436–474.
- Fama, E.F. 1984. Forward and spot exchange rates. *Journal of Monetary Economics* 14: 319–338.
- Frankel, J.A. 1979. On the mark: A theory of floating exchange rates based on real interest differential. *American Economic Review* 69: 610–622.
- Hakkio, C.S., and A. Sibert. 1995. The foreign exchange risk premium: Is it real? *Journal of Money, Credit and Banking* 27: 301–317.
- Hansen, L.P., and R.J. Hodrick. 1980. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. *Journal of Political Economy* 88: 829–853.
- Hodrick, R.J. 1987. *The empirical evidence on the efficiency of forward and futures foreign exchange markets*. New York: Harwood Academic.
- Jiang, C., and T.C. Chiang. 2000. Do foreign exchange risk premiums relate to the volatility in the foreign exchange and equity markets? *Applied Financial Economics* 10: 95–104.
- Korajczyk, R.A. 1985. The pricing of forward contracts for foreign exchange. *Journal of Political Economy* 93: 346–368.
- Korajczyk, R.A., and C.J. Viallet. 1992. Equity risk premia and the pricing of foreign exchange risk. *Journal of International Economics* 33: 199–220.
- Levine, R. 1989. The pricing of forward exchange rates. *Journal of International Money and Finance* 8: 163–180.
- Nelson, D. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59: 347–370.
- Newey, W.K., and K.D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55: 703–708.
- Shiller, R. 2014. Speculative asset prices. *American Economic Review* 104: 1486–1517.
- Trask, J. 2017. Geopolitical risks and impacts on currency markets. *The Week*, November 17. Available at: <https://www.theweek.co.uk/89705/geopolitical-risks-and-impacts-on-currency-markets>

Treasury Inflation-Protected Securities

8

Quentin C. Chu and Deborah N. Pittman

Abstract

In January 1997, the U.S. Treasury began to issue inflation-protected securities (TIPS). TIPS protect investors from inflation by linking the principal and coupon payments to the Consumer Price Index (CPI). Empirical studies of TIPS have focused on their term structure, their role in diversifying portfolios, and their usefulness in generating a measurement of expected inflation. This chapter discusses TIPS unique characteristics, the role they play in aggregating inflation information and price discovery in Treasury security markets.

An econometric method is proposed to identify the speed and timing of TIPS price adjustments to inflation information. The econometric method is based on a pooled time-series cross-sectional regression analysis of TIPS daily holding period returns on inflation surprises. The inflation surprise is measured by the difference between actual inflation and the observed nominal and real interest rate spread. The speed and timing of TIPS price adjustments are revealed in the estimated cumulative regression coefficients. In addition, vector error correction model and common-factor model are applied to investigation price discovery in Treasury bond and TIPS markets.

Keywords

Inflation information aggregation · Inflation risk premium · Information risk · Price discovery · Real interest rate · Treasury inflation-protected securities

Q. C. Chu (✉)
University of Memphis, Memphis, TN, USA
e-mail: qchu@memphis.edu

D. N. Pittman
Rhodes College, Memphis, Tennessee, USA

8.1 Introduction

The fundamental notion behind inflation protection is to preserve the purchasing power of money. By linking value to the Consumer Price Index (CPI), Treasury inflation-protected securities (TIPS) provide investors with a “real” rate of return. This security can be viewed as one of the safest financial assets due to its minimal exposure to default risk and uncertain inflation. Today, inflation protection may be accomplished by linking investment principal to some form of a price index, such as the Consumer Price Index (CPI) in the United States, Canada, the United Kingdom, and Iceland; the Wholesale Price Index (WPI) in Finland, Brazil, and Argentina; and equities and gold in France. In essence, investors purchasing inflation-protected securities are storing a basket of goods for future consumption. Fifteen countries, including the U.S., have issued inflation-protected securities, starting in the 1940s.¹ Some of the countries had extremely high inflation, such as Mexico and Brazil (114.8% and 69.2% in the year prior to the introduction of inflation-protected securities), and others had moderate inflation like Sweden and New Zealand (4.4% and 2.8%). As of the end of 2008, inflation-protected securities made up 24% of the United Kingdom’s total outstanding debt portfolio, 15% of France’s outstanding debt portfolio, and 10% of the United States’ outstanding debt portfolio (U.S. GAO 2009).

8.2 Size of Market

The first U.S. Treasury inflation-protected securities were issued in January 1997. Thirty-two issues of TIPS are traded in the U.S. market as of October, 2010, with maturities ranging from 2011 to 2040. According to the Treasury Department, TIPS are an important component of Treasury’s debt management strategy, although they represent by June 2009 only about 7% of the U.S. Treasury debt that is outstanding. Over the past year, Treasury has increased overall TIPS issuance by a relatively small amount, and replaced the 20-year TIPS with 30-year TIPS. As of June 2009 the U.S. Treasury had \$3.4 trillion outstanding in Treasury notes, \$1.6 trillion outstanding in Treasury bills, \$0.6 trillion in Treasury bonds, and \$0.5 trillion in TIPS. Between December 2007 and June 2009, the Treasury issued \$1.6 trillion in new debt. TIPS only increased by \$0.06 trillion during that period of time, representing only 3.7% of the new debt (U.S. GAO 2009).

¹ According to the date of introduction of inflation-protected securities, these countries are: Finland, France, Sweden, Israel, Iceland, Brazil, Chile, Colombia, Argentina, the U.K., Australia, Mexico, Canada, New Zealand, and the U.S.

8.3 Reference CPI

TIPS provide two guarantees: (1) that investors will receive an inflation-adjusted amount or the real par value at maturity, whichever is greater; and (2) that coupon payments will be adjusted for inflation occurring between issuance and payment dates. The consumer price index for all urban consumers (CPI-U) was chosen to measure price level changes. As the most widely used index, the CPI-U is generally accepted as a measure of inflation. The Bureau of Labor Statistics (BLS) surveys prices each month. Around the middle of the month, it announces changes in retail prices experienced by American consumers during the previous month. Although the CPI-U is not a cost of living index, it is the deflator commonly used to adjust wages and salaries for collective bargaining agreements, and to keep pensions, rents, and child support payments in line with inflation.

8.4 Conversion from Real to Nominal Prices

Prices quotes in the market are real clean prices. When there is a transaction, the invoice price is computed from the real clean price, the reference CPI-U index, and the accrued nominal coupon. The coupon payments and the lump sum payment at maturity are adjusted according to inflation rates. With a fixed real coupon rate, the adjustment to a nominal coupon payment is accomplished by multiplying the principal value by one plus the inflation rate between the issuance date and the coupon payment date. Inflation-protected securities set a floor (par value), an implicit put option, guaranteeing the bond's value on maturity date will not fall below its face value if the U.S. experiences cumulative deflation during the entire life of the TIPS, a highly unlikely event.

8.5 Three-Month Lag Effect

One feature of the TIPS that impedes its use as a perfect measure of the *ex ante* real rate is the CPI indexing procedure. There is a 3-month lag in the CPI indexing system for TIPS. Figure 1 indicates how the reference CPI is calculated on May 15, 2007. The reference CPI for May 1, 2007, is the CPI-U for the third-previous calendar month, i.e., the announced CPI for February 2007. The Bureau of Labor Statistics surveys price information for the February CPI during the month of February, and then announces the February CPI on March 16, 2007. The reference CPI for any other day of May is calculated by linear interpolation between the CPIs of February and March (the CPI for March became available on April 17, 2007). Once the March CPI is announced, the reference CPI for any day in May 2007 is known. The reference CPI for May 15, 2007 can be calculated according to the following formula:

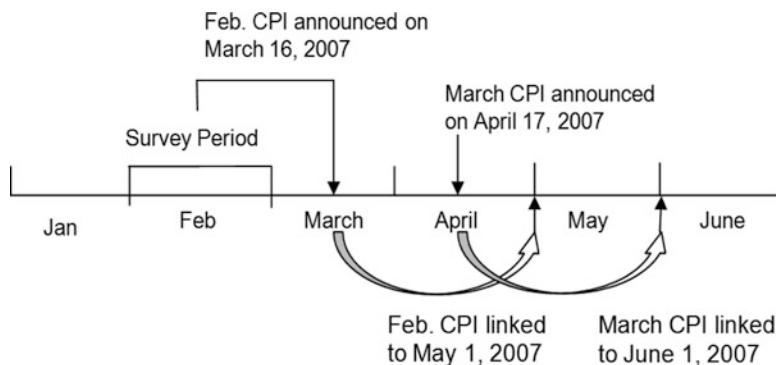


Fig. 1 Calculation of Reference CPI This figure illustrates the lag effect in indexing the CPI. Due to CPI-U reporting procedures, the reference CPI for May 1, 2007, is linked to the February CPI-U, and the reference CPI for June 1, 2007, is linked to the March CPI-U

$$\begin{aligned}
 RCPI_{May15} &= CPI_{Feb} + (14/31) (CPI_{March} - CPI_{Feb}) \\
 &= 203.499 + (14/31) (205.352 - 203.499) = 204.33584
 \end{aligned}$$

where $RCPI$ represents the reference CPI for a particular day.²

The principal value of TIPS on any particular trading day is determined by multiplying the face value at the issuance by an applicable index ratio. The index ratio is defined as the reference CPI applicable to the trading date divided by the reference CPI applicable to the original issuing date.

8.6 Public Issuance

The design of the U.S. TIPS underwent considerable discussion in determining the linking price index, the cash flow structure, the optimal length of maturity, the auction mechanism, and the amount of issuance. TIPS are auctioned through the Dutch uniform price auction method used by other Treasury securities. Participants submit bids in terms of real yields. The highest accepted yield is used to price the newly issued TIPS for all participants (Roll 1996). TIPS are eligible for stripping into their principal and interest components in the Treasury's Separate Trading of Registered Interest and Principal of Securities Program. Since March 1999, the U.S. Treasury Department has allowed all TIPS interest components with the same maturity date to be interchangeable (fungible). Fungibility is designed to improve the liquidity of stripped interest components of TIPS, and hence increase demand for the underlying inflation-protected securities. Other Treasury securities are strippable

²The U.S. Treasury website (<http://www.publicdebtreas.gov>) posts the reference CPI for the following month after the CPI announcement date which is around the 15th of each month.

as well. Each issue has a unique CUSIP number for identification purposes, which is also used in a case of reopening (Grievess and Sunner1999).

8.7 Tax Disadvantage: Phantom Income

One disadvantage of TIPS is the potential tax liability on phantom income. Although the securities are exempt from state and local taxes, they are subject to federal taxation. Positive accrued inflation compensation, if any, is reportable income, even though the inflation-adjusted principal will not be received until maturity. Some taxable investors may thus hesitate to invest in TIPS, while others with non-taxable accounts such as retirement accounts might find this market attractive. Consequently, investor tax brackets may affect decisions about including TIPS in a portfolio. The emergence of mutual funds and exchange traded funds (ETFs) specializing in TIPS have attracted more individual investment in the form of IRA and 401(K) savings, although these investors are more likely to buy and hold.

8.8 TIPS as an Asset Class

The real yield of 10-year TIPS averaged 2.06% in 2003, 1.83% in 2004, 1.81% in 2005, 2.31% in 2006, 2.29% in 2007, 1.77% in 2008, and 1.66% in 2009. Comparable yields of conventional 10-year Treasury bonds were 4.10% (2003), 4.27% (2004), 4.29% (2005), 4.80% (2006), 4.63% (2007), 3.66% (2008), and 3.26% (2009).

The performance of TIPS as an asset class has been extremely well since its inception in 1997. The Sharpe ratio for a 10-year TIPS has averaged 0.61 from 1997 to 2009, while the Sharpe ratio for a 10-year Treasury bond has averaged 0.58 for the same period. This compares to 0.01 for the S&P 500 (Katz and Palazzolo 2010).

Inflation-protected securities offer an alternative financial vehicle for portfolio management. Since the returns on nominal bonds are fixed in nominal terms, they provide no hedge against uncertain inflation. Kaul (1987) and Chu et al. (1995) have documented a negative correlation between equity returns and inflation in the U.K. That is, investors in equity markets suffer during periods of unexpected high inflation. Inflation-protected securities, by linking returns to the movement of a price index, provide a hedge for investors who have a low risk tolerance for unexpected inflation. Those most averse to inflation will purchase inflation-protected securities, and those less sensitive to inflation will purchase the riskier nominal bond. These unique U.S. Treasury instruments that protect against future inflation are now viewed as belonging in most well-diversified investment portfolios (Roll2004). The Investment Company Institute survey data finds that mutual funds invested only in TIPS increased from \$11.8 billion in December 2002 to \$54.4 billion in May 2009 (U.S. GAO 2009).

8.9 Size of the Inflation Risk Premium

According to the Employment Act of 1946, one of the four primary goals of the U.S. federal government economic policy is to stabilize prices through a low inflation rate. Inflation-protected securities provide a way for the public to evaluate the government's performance in controlling inflation. For a constant level of expected inflation, the wider the yield spread between nominal and real bonds, the higher the inflation risk premium, and presumably the lower the public's confidence in the monetary authorities. When there is no risk of inflation, the inflation risk premium is reduced, if not eliminated completely. Benninga and Protopapadakis (1983) revise the Fisher equation to incorporate an inflation risk premium. Prior to the issuance of the U.S. TIPS, Chu et al. (1995) measured the inflation risk premium in the United Kingdom between 1985 and 1991 using the real yield on the indexed linked gilt (similar to the U.S. TIPS), and found it was 2.41% and statistically significant. Recent research has found that the inflation risk premium in the United States over the past decade has been insignificant and less than 25 basis points (Dudley et al. 2009).

8.10 Cost of TIPS to the Treasury Since Inception

One motivation for the issuance of inflation-protected securities is that governments can reduce public financing costs through reducing the interest paid on public debt by the amount of an inflation risk premium. Rates on Treasury securities are thought to consist of the expected real rate, plus expected inflation and an inflation risk premium, while TIPS provide investors with a real rate of interest plus actual inflation. The TIPS real return is guaranteed, whatever the course of inflation. The lower the *ex post* inflation in the U.S., the lower the amount of interest paid on the outstanding TIPS debt, while the interest on conventional bonds remains constant at the contracted amount when issued.

Market experts have measured costs of TIPS versus cost of conventional bonds by comparing the cost of TIPS and nominal securities based on what inflation actually was during a given time period. This approach concludes that the TIPS program has been less cost-effective for Treasury than nominal securities. A 2008 study by a Federal Reserve economist estimated that the total cost of the TIPS program through October 2007 was between \$4.5 billion and \$7.5 billion (U.S. GAO 2009). Through July 2009, Treasury's Office of Debt Management (ODM) estimated the total cost differential of TIPS at \$10 billion to date. This more recent study includes the anomalous period characterized by the financial market crisis and the related flight to quality which made nominal securities relatively inexpensive for Treasury.

However, economists generally agree that at least part of the relatively higher cost of the TIPS program is due to its start-up costs that included a liquidity premium since TIPS were traded off the run. Market experts and the major institutional investors had indicated that other factors that may have contributed to the relatively

higher cost to Treasury of the TIPS program to date are investors' perceptions of the Treasury's lack of commitment to the program and the relatively low inflation experienced throughout the life of the program. Since the TIPS program was introduced, the annual Consumer Price Index for All Urban Consumers (CPI-U) percent change never exceeded 3.80%, which is low by historical standards. The CPI-U averaged 7.08% and 5.55% during the 1970s and 1980s, respectively (U.S. GAO 2009).

8.11 Liquidity Premiums

The surprisingly high yields of TIPS, relative to conventional Treasuries, have been both a puzzle and a concern to the U.S. Treasury Department due to the excess interest cost. Sack (2000), Shen and Corning (2001) and Sack and Elsasser (2004) focus on the possible reasons for a "too high" TIPS real yield and consider a liquidity premium, a lower supply, and a lack of knowledge about the new TIPS as possible reasons. Sack and Elsasser (2004) conclude, liquidity has improved over the past 7 years, the supply has increased, and there is wider acceptance of the securities, yet the relatively high TIPS real yield remains. D'Amico et al. (2008) estimate that the liquidity premium of the TIPS was as large as 200 basis points during its early years of trading, but had trended down to 50 basis points for 5-year maturities after 2004. While the prices of both nominal securities and TIPS could be influenced by inflation risk premiums and liquidity premiums, D'Amico et al. (2008) found these premiums to be relatively small and stable after 2004.

8.12 Observable Expected Real Rate

TIPS provide a direct measure of expected real interest rates that may help policy makers make economic decisions. According to economic theory, most savings, consumption and investment decisions depend critically on the expected real rate of interest, the interest rate one earns after adjusting the nominal interest rate for the expected rate of inflation. Real interest rates measure the real growth rate of the economy and the supply and demand for capital in the market. Before the trading of inflation-protected securities, there was no security in the U.S. offering coupon and principal payments linked to inflation and therefore enabling measurement of the expected real rate. Empirical studies testing the relationship between expected real rates and other macroeconomic variables have relied instead on indirect measures of the expected real rate such as *ex post* real rates estimated by subtracting actual inflation from realized nominal holding-period returns (Pennachi 1991). Inflation-protected securities permit the direct study of the real interest rate. Wilcox (1998) includes this as one benefit motivating the Treasury to issue these new securities. For example, Chu et al. (2003) used measurements of the *ex ante* real rate with a constant 10-year maturity derived from TIPS prices, and found that a cointegrated system

exists between the real and nominal rates. This casts doubt on the accuracy of previous Fisher Effect testing that assumed a constant or stationary real rate.

8.13 Information Content of Maturing TIPS

On July 15, 2002, the first maturity of a TIPS issue occurred. The distinctive properties of maturing TIPS during the last 6-month coupon reveal the market's anticipated inflation regime (Chu et al. 2004). During their last coupon period, TIPS have characteristics of both nominal bonds and real bonds. After the next-to-last coupon payment date on January 15, 2002, the July 2002 issue of TIPS had a single cash flow remaining. The distinctive properties of a maturing issue result in a special environment that permit researchers to observe relatively clean and useful information about an anticipated inflation regime. With a single cash flow remaining, there is no need to set the reinvestment rate for periodic cash flows in the pricing of maturing TIPS. Moreover, there is no difference in tax treatment to anyone purchasing Treasury bills or TIPS during the last 6 months prior to expiration, and there is no call feature. The relative pricing between the maturing TIPS and the synthetic Treasury bill reveals the market anticipation of an inflationary or a disinflationary regime.

As it has only one role to play, the pricing of a maturing TIPS represents a pure successive forecast of the target CPI linked to the maturity date of the TIPS. Under the assumption of risk-neutrality, the sequence of forecasts implied in maturing TIPS prices can be modeled as a martingale (Chu et al. 2007).

As the forecasted CPI must converge to the actual CPI to avoid any possibility of an arbitrage profit, a consistent upward or downward movement of the CPI forecasts implied in maturing TIPS prices indicates inflation hedge behavior. A consistent upward movement in the forecast series would indicate the market anticipates a disinflationary regime, while a consistent downward movement would indicate an anticipated inflationary regime.

Most inflation outlooks are based on survey data or expert opinions. Few studies provide empirical evidence derived from security prices to reveal general overriding concerns about disinflation in the U.S. market. There is evidence that TIPS prices during a maturing issue's last coupon period are able to discern whether the market is anticipating a disinflationary regime or an inflationary regime. Monetary authorities can use this timely identification of the market's concern about a particular regime in adjusting monetary policy.

8.14 Inflation Information Aggregation

A nominal Treasury security alone does not allow measurement of the aggregation of information about inflation as it occurs. It rather incorporates estimates of future inflation information in its expected yield. The price of the TIPS, however, responds in a different manner. Its nominal price will increase over its life as the CPI level

changes, because all future coupons and the principal adjust automatically to the CPI level. TIPS prices respond to changes in inflation as they occur; the higher the price index, the higher the nominal income from the bond.

Offering a certain real return, TIPS hedge against a rising price level. They are uniquely structured to aggregate inflation information prior to the monthly public announcement, due to the direct link between the future cash flows and *ex post* CPI. Since the cash flows associated with TIPS depend on actual inflation and a contractual real return, TIPS prices react far differently over time than conventional bond prices. While conventional bond prices respond to changes in the expected rate of inflation until maturity and also to changes in the expected real rate, the TIPS prices respond only to changes in actual inflation and changes in the expected real rate, assuming contemporaneous adjustment of the contractual cash flow to the current CPI.

Chu et al. (2011), using pooled time-series cross-section data from three matured TIPS issues, find that TIPS prices respond to the monthly update of the CPI, efficiently aggregating near-term inflation information. The evidence supports a market that is highly informed about upcoming inflation starting 44 business days prior to a CPI announcement date. In fact, 29% of the cumulative adjustment to information about the upcoming month's inflation is already incorporated into the TIPS price before the survey period begins. Moreover, the cumulative effects of unexpected inflation on TIPS returns peaks on the last day of the month as the survey period ends, with 98% of the inflation adjustment already in the TIPS price. After the month has ended, there is a statistically insignificant reversing trend that persists during the compilation period until the day before the public announcement. A significant adjustment on the announcement day returns the cumulative adjustment to a level slightly higher than at the end of the month. Thus, the market is very efficient in monitoring and responding to changes in consumer prices as they occur.

When the monthly CPI survey published by the *Blue Chip Financial Forecasts* was used to measure expected inflation instead of the breakeven inflation rate, there was no significant difference in the timing of TIPS price adjustments. This provides evidence that the market-determined measure of expected inflation, even using securities with a 5-year maturity, is robust in capturing near-term inflation surprises.

8.15 Price Discovery and Information Risk

Both nominal Treasury and TIPS markets share a common component – the expected real return – but they contain other return components that behave very differently in the presence of inflation, and therefore attract different clienteles who must hold non-replicating portfolios. Investors holding TIPS are attracted to its inflation hedging properties. However, they cannot diversify into other securities without giving up the inflation hedging properties. If there is information risk by holding the TIPS, because the real rate information is likely to be revealed first in the nominal Treasury market, they will need additional compensation for taking this risk. Moreover, changes in real rates of interest create a more volatile price for TIPS

than for nominal Treasuries due to longer real durations for TIPS (Roll 2004), so perception of an informational disadvantage is a consequential risk.

TIPS investors are at a disadvantage in price discovery with a 1-day lag (Chu et al. 2005). While it is possible that informed traders could transact their trades in either market, the evidence is that information flows unilaterally from the Treasury bond market to the TIPS market. The reason for such a lag in price innovation is also analogous to Easley and O'Hara's (2004) description of stocks with more public information having a "greater institutional following." Bond market participants who have private information about the real rate of interest could act on this information either in the nominal Treasury market or in the TIPS market. Chu et al. (2005) find that they choose to act in the former. This is likely because the nominal Treasury market, with its highly developed infrastructure and large number of traders in both the spot and derivative markets, is the most profitable vehicle for trading on new information about expected real interest rates. The trading aggregates new information into the price, making information public.

8.16 Conclusion

Since the trading of TIPS in January 1997, the pooled time-series cross-sectional TIPS price data are used to investigate how quickly TIPS prices respond to the monthly update of the CPI. Our empirical results based on the proposed econometric method show that TIPS adjust to inflation information without delay during the U.S. Consumer Price Index (CPI) survey period and even before the beginning of the survey period. The cumulative effect of inflation surprise on the TIPS holding period returns peaks at the end of the survey period. The TIPS data are also used to investigate the presence of information risk in two closely linked interest rate securities traded in separate markets: the Treasury nominal bond market and the TIPS market. We find that information flows unilaterally from the Treasury bond market to the TIPS market with a 1-day lag. The information risk arising from asymmetric information flows may cause less informed traders to demand a higher rate of return.

References

- Benninga, Simon, and Aris Protopapadakis. 1983. Nominal and real interest rates under uncertainty: The Fisher theorem and the term structure. *The Journal of Political Economy* 91: 856–867.
- Chu, Quentin C., Cheng F. Lee, and Deborah N. Pittman. 1995. On the inflation risk premium. *Journal of Business Finance and Accounting* 22: 881–892.
- Chu, Quentin C., Deborah N. Pittman, and Q.Yu. Linda. 2003. Real rates, nominal rates, and the Fisherian link. *International Review of Financial Analysis* 12: 189–205.
- . 2004. Information content of maturing TIIS. *The Journal of Fixed Income* 13: 90–99.
- . 2005. Information risk in TIPS market: An analysis of nominal and real interest rates. *Review of Quantitative Finance and Accounting* 24: 235–250.

- Chu, Quentin C., Deborah N. Pittman, and Jeng-Hong Chen. 2007. Inflation or disinflation? Evidence from maturing US Treasury Inflation-Protected Securities. *Applied Economics* 39: 361–372.
- Chu, Quentin C., Deborah N. Pittman, and Linda Q. Yu. 2011. When do TIPS prices adjust to information? *Financial Analysts Journal* 67: 59–73.
- D’Amico, Stefania, Don H. Kim, and Min Wei. 2008. Tips from TIPS: The informational content of treasury inflation-protected security prices. *Bank for International Settlements working papers*, No. 248.
- Dudley, William C., Jennifer Roush, and Michelle Steinberg Ezer. 2009. The case for TIPS: An examination of the costs and benefits. *Economic Policy Review*, Federal Reserve Bank of New York. 15: 1–15.
- Easley, David, and Maureen O’Hara. 2004. Information and the cost of capital. *Journal of Finance* 59: 1553–1583.
- Grieses, Robin, and Michael W. Sunner. 1999. Fungible STRIPS for the U.S. Treasury’s inflation-indexed securities. *The Journal of Fixed Income* 9: 55–62.
- Katz, Michael, and Christopher Palazzolo. 2010. *Inflation in 2010 and beyond? Practical considerations for institutional asset allocation*. Greenwich: AQR Capital Management, LLC.
- Kaul, Gautam. 1987. Stock returns and inflation: The role of the monetary sector. *Journal of Financial Economics* 18: 253–276.
- Pennachi, George G. 1991. Identifying the dynamics of real interest rates and inflation: Evidence using survey data. *Review of Financial Studies* 4: 53–86.
- Roll, Richard. 1996. U.S. Treasury inflation-protected bonds: The design of a new security. *The Journal of Fixed Income* 6: 9–28.
- . 2004. Empirical TIPS. *Financial Analysts Journal* 60: 31–53.
- Sack, Brian. 2000. Deriving inflation expectations from nominal inflation-indexed treasury yields. *The Journal of Fixed Income* 10: 6–17.
- Sack, Brian, and Robert Elsasser. 2004. Treasury inflation-indexed debt: A review of the U.S. experience. *Economic Policy Review*, Federal Reserve Bank of New York 10: 47–63.
- Shen, Paul, and Jonathan Corning. 2001. Can TIPS help identify long-term inflation expectations? *Economic Review*, Federal Reserve Bank of Kansas City. 86: 61–87.
- United States General Accounting Office. 2009. Treasury inflation protection securities should play a heightened role in addressing debt management changes. GAO 09–932.
- Wilcox, David W. 1998. The introduction of indexed government debt in the United States. *Journal of Economic Perspectives* 12: 219–227.



Wayne E. Ferson

Abstract

The asset pricing models of financial economics describe the prices and expected rates of return of securities based on arbitrage or equilibrium theories. These models are reviewed from an empirical perspective, emphasizing the relationships among the various models.

Keywords

Arbitrage · Arbitrage pricing model · Beta pricing model · Capital asset pricing model · Consumption · Durable goods · Factor models · Financial assets · Habit persistence · Intertemporal marginal rate of substitution · Mean variance efficiency · Portfolio optimization · Risk aversion · Stochastic discount factor · Systematic risk

Asset pricing models describe the prices or expected rates of return of financial assets, which are claims traded in financial markets. Examples of financial assets are common stocks, bonds, options, and futures contracts. The asset pricing models of financial economics are based on two central concepts. The first is the “no arbitrage principle,” which states that market forces tend to align the prices of financial assets so as to eliminate arbitrage opportunities. An arbitrage opportunity arises if assets can be combined in a portfolio with zero cost, no chance of a loss, and a positive probability of gain. Arbitrage opportunities tend to be eliminated in financial markets because prices adjust as investors attempt to trade to exploit the arbitrage opportunity. For example, if there is an arbitrage opportunity because the price of security A is too low, then traders’ efforts to purchase security A will tend to drive up its price, which will tend to eliminate

W. E. Ferson (✉)
University of Southern California, Los Angeles, CA, USA
e-mail: ferson@marshall.usc.edu

the arbitrage opportunity. The arbitrage pricing model (APT), (Ross1976) is a well-known asset pricing model based on arbitrage principles.

The second central concept in asset pricing is “financial market equilibrium.” Investors’ desired holdings of financial assets are derived from an optimization problem. A necessary condition for financial market equilibrium in a market with no frictions is that the first-order conditions of the investor’s optimization problem are satisfied. This requires that investors are indifferent at the margin to small changes in their asset holdings. Equilibrium asset pricing models follow from the first-order conditions for the investors’ portfolio choice problem, and a market-clearing condition. The market-clearing condition states that the aggregate of investors’ desired asset holdings must equal the aggregate “market portfolio” of securities in supply.

Differences among the various asset pricing models arise from differences in their assumptions about investors’ preferences, endowments, production and information sets, the process governing the arrival of news in the financial markets, and the types of frictions in the markets. Recently, models have been developed that emphasize the role of human imperfections in this process. For a review of this “behavioral finance” perspective, see Barberis and Shleifer (2003).

Virtually all asset pricing models are special cases of the fundamental equation:

$$P_t = E_t\{m_{t+1}(P_{t+1} + D_{t+1})\}, \quad (1)$$

where P_t is the price of the asset at time t and D_{t+1} is the amount of any dividends, interest or other payments received at time $t + 1$. The market wide random variable m_{t+1} is the “stochastic discount factor” (SDF). By recursive substitution in Eq. 1, the future price may be eliminated to express the current price as a function of the future cash flows and SDFs only: $P_t = E_t\{\sum_{j=0}^{\infty} (\prod_{k=1}^j m_{t+k}) D_{t+j}\}$. Prices are obtained by “discounting” the payoffs, or multiplying by SDFs, so that the expected “present value” of the payoff is equal to the price.

We say that a SDF “prices” the assets if Eq. 1 is satisfied. Any particular asset pricing model may be viewed simply as a specification for the stochastic discount factor. The random variable m_{t+1} is also known as the benchmark pricing variable, equivalent martingale measure, Radon–Nicolom derivative, or intertemporal marginal rate of substitution, depending on the context. The representation in Eq. 1 goes at least back to Beja (1971), while the term “stochastic discount factor” is usually ascribed to Hansen and Richard (1987).

Assuming nonzero prices, Eq. 1 is equivalent to:

$$E_t(m_{t+1}\mathbf{R}_{t+1} - \mathbf{1}) = 0, \quad (2)$$

where $\mathbf{R}_{t+1}$ is the vector of primitive asset gross returns and $\mathbf{1}$ is an N -vector of ones. The gross return $R_{i,t+1}$ is defined as $(P_{i,t+1} + D_{i,t+1})/P_{i,t}$, where $P_{i,t}$ is the price of the asset i at time t and $D_{i,t+1}$ is the payment received at time $t + 1$. Empirical tests of asset pricing models often work directly with asset returns in Eq. 2 and the relevant definition of m_{t+1} .

Without more structure the Eqs. 1 and 2 have no content, because it is always possible to find a random variable m_{t+1} for which the equations hold. There will be some m_{t+1} that “works,” in this sense, as long as there are no redundant asset returns. For example, take a sample of asset gross returns with a nonsingular covariance matrix and let m_{t+1} be, $[1'(E_t\{\mathbf{R}_{t+1}\mathbf{R}_{t+1}'\}) - 1]\mathbf{R}_{t+1}$. Substitution in to Eq. 2 shows that this SDF will always “work” in any sample of returns. The ability to construct an SDF as a function of the returns that prices all of the included assets, is essentially equivalent to the ability to construct a minimum-variance efficient portfolio and use in as the “factor” in a beta pricing model, as described below.

With the restriction that m_{t+1} is a strictly positive random variable, Eq. 1 becomes equivalent to the no arbitrage principle, which says that all portfolios of assets with payoffs that can never be negative but are positive with positive probability, must have positive prices (Beja 1971; Rubinstein 1976; Ross 1977; Harrison and Kreps 1979; Hansen and Richard 1987.)

While the no arbitrage principle places restrictions on m_{t+1} , empirical work more typically explores the implications of equilibrium models for the SDF based on investor optimization. A representative consumer–investor’s optimization implies the Bellman equation:

$$J(W_t, s_t) \equiv \max E_t\{U(C_t, \cdot) + J(W_{t+1}, s_{t+1})\}, \quad (3)$$

where $U(C_t, \cdot)$ is the utility of consumption expenditures at time t , and $J(\cdot)$ is the indirect utility of wealth. The notation allows that the direct utility of current consumption expenditures may depend on other variables such as past consumption expenditures or the current state variables. The state variables, s_{t+1} , are sufficient statistics, given wealth, for the utility of future wealth in an optimal consumption–investment plan. Thus, the state variables represent future consumption–investment opportunity risk. The budget constraint is: $W_{t+1} = (W_t - C_t)\mathbf{x}'\mathbf{R}_{t+1}$, where $\mathbf{x}$ is the portfolio weight vector, subject to $\mathbf{x}'\mathbf{1} = 1$.

If the allocation of resources to consumption and investment assets is optimal, it is not possible to obtain higher utility by changing the allocation. Suppose an investor considers reducing consumption at time t to purchase more of (any) asset. The expected utility cost at time t of the foregone consumption is the expected product of the marginal utility of consumption expenditures, $U_c(C_t, \cdot) > 0$ (where a subscript denotes partial derivative), multiplied by the price of the asset, and which is measured in the same units as the consumption expenditures. The expected utility gain of selling the investment asset and consuming the proceeds at time $t + 1$ is $E_t\{(P_{i,t+1} + D_{i,t+1})J_w(W_{t+1}, s_{t+1})\}$. If the allocation maximizes expected utility, the following must hold: $P_{i,t}E_t\{U_c(C_t, \cdot)\} = E_t\{(P_{i,t+1} + D_{i,t+1})J_w(W_{t+1}, s_{t+1})\}$ which is equivalent to Eq. 1, with

$$m_{t+1} = \frac{J_w(W_{t+1}, s_{t+1})}{E_t\{U_c(C_t, \cdot)\}}. \quad (4)$$

The m_{t+1} in Eq. 4 is the “intertemporal marginal rate of substitution” (IMRS) of the consumer–investor.

Asset pricing models typically focus on the relation of security returns to aggregate quantities. It is therefore necessary to aggregate the first-order conditions of individuals to obtain equilibrium expressions in terms of aggregate quantities. Then, Eq. 4 may be considered to hold for a representative investor who holds all the securities and consumes the aggregate quantities. Theoretical conditions that justify the use of aggregate quantities are discussed by Gorman (1953), Wilson (1968), Rubinstein (1974), and Constantinides (1982), among others. When these conditions fail, investors' heterogeneity will affect the form of the asset pricing relation. The effects of heterogeneity are examined by Lintner (1965), Brennan and Kraus (1978), Lee et al. (1990), Constantinides and Duffie (1996), and Sarkissian (2003), among others.

Typically, empirical work in asset pricing focuses on expressions for expected returns and excess rates of return. The expected excess returns are modeled in relation to the risk factors that create variation in m_{t+1} . Consider any asset return $R_{i,t+1}$ and a reference asset return, $R_{0,t+1}$. Define the excess return of asset i , relative to the reference asset as $r_{i,t+1} = R_{i,t+1} - R_{0,t+1}$. If Eq. 2 holds for both assets it implies:

$$E_t\{m_{t+1}r_{i,t+1}\} = 0 \text{ for all } i. \quad (5)$$

Use the definition of covariance to expand Eq. 5 into the product of expectations plus the covariance, obtaining:

$$E_t\{r_{i,t+1}\} = \frac{\text{Cov}_t(r_{i,t+1}; -m_{t+1})}{E_t\{m_{t+1}\}}, \text{ for all } i, \quad (6)$$

where $\text{Cov}_t(\cdot, \cdot)$ is the conditional covariance. Equation 6 is a general expression for the expected excess return from which most of the expressions in the literature can be derived.

Equation 6 implies that the covariance of return with m_{t+1} , is a general measure of "systematic risk." This risk is systematic in the sense that any fluctuations in the asset return that are uncorrelated with fluctuations in the SDF are not "priced," meaning that these fluctuations do not command a risk premium. For example, in the conditional regression $r_{it+1} = a_{it} + b_{it} m_{t+1} + u_{it+1}$, then $\text{Cov}_t(u_{it+1}, m_{t+1}) = 0$. Only the part of the variance in a risky asset return that is correlated with the SDF is priced as risk.

Equation 6 displays that a security will earn a positive risk premium if its return is negatively correlated with the SDF. When the SDF is an aggregate IMRS, negative correlation means that the asset is likely to return more than expected when the marginal utility in the future period is low, and less than expected when the marginal utility and the value of the payoffs, is high. For a given expected payoff, the more negative the covariance of the asset's payoffs with the IMRS, the less desirable the distribution of the random return, the lower the value of the asset and the larger the expected compensation for holding the asset given the lower price.

9.1 The Capital Asset Pricing Model

One of the first equilibrium asset pricing models was the Capital Asset Pricing Model (CAPM), developed by Sharpe (1964), Lintner (1965), and Mossin (1966). The CAPM remains one of the foundations of financial economics, and a huge number of theoretical papers refine the assumptions and provide derivations of the CAPM. The CAPM states that expected asset returns are given by a linear function of the assets' "betas," which are their regression coefficients against the market portfolio. Let R_{mt} denote the gross return for the market portfolio of all assets in the economy. Then, according to the CAPM,

$$E(R_{it+1}) = \delta_0 + \delta_1 \beta_i \quad (7)$$

where $\beta_i = \text{Cov}(R_i, R_m) / \text{Var}(R_m)$.

In Eq. 7, $\delta_0 = E(R_{0t+1})$, where the return R_{0t+1} is referred to as a "zero-beta asset" to R_{mt+1} because the condition $\text{Cov}(R_{0t+1}, R_{mt+1}) = 0$.

To derive the CAPM, it is simplest to assume that the investor's objective function in Eq. 3 is quadratic, so that $J(W_{t+1}, S_{t+1}) = V\{E_t(R_{pt+1}), \text{Var}_t(R_{pt+1})\}$ where R_{pt+1} is the investor's optimal portfolio. The function $V(\cdot, \cdot)$ is increasing in its first argument and decreasing in the second if investors are risk averse. In this case, the SDF of Eq. 4 specializes as: $m_{t+1} = a_t + b_t R_{pt+1}$. In equilibrium, the representative agent must hold the market portfolio, so $R_{pt+1} = R_{mt+1}$. Equation 7 then follows from Eq. 6, with this substitution.

9.2 Consumption-Based Asset Pricing Models

Consumption models may be derived from Eq. 4 by exploiting the envelope condition, $U_c(\cdot) = J_w(\cdot)$, which states that the marginal utility of current consumption must be equal to the marginal utility of current wealth, if the consumer has optimized the tradeoff between the amount consumed and the amount invested.

Breeden (1979) derived a consumption-based asset pricing model in continuous time, assuming that the preferences are time-additive. The utility function for the lifetime stream of consumption is $\sum_t \beta^t U(C_t)$, where β is a time preference parameter and $U(\cdot)$ is increasing and concave in current consumption, C_t . Breeden's model is a linearization of Eq. 1, which follows from the assumption that asset values and consumption follow diffusion processes (Bhattacharya 1981; Grossman and Shiller 1982). A discrete-time version follows Lucas (1978), assuming a power utility function:

$$U(C) = [C^{1-\alpha} - 1] / (1 - \alpha) \quad (8)$$

where $\alpha > 0$ is the concavity parameter of the period utility function. This function displays constant relative risk aversion equal to α . "Relative risk aversion" in consumption is defined as: $Cu''(C)/u'(C)$. Absolute risk aversion is defined as: u''

$(C)/u'(C)$. Ferson (1983) studied a consumption-based asset pricing model with constant absolute risk aversion.

Using Eq. 8 and the envelope condition, the IMRS in Eq. 4 becomes:

$$m_{t+1} = \beta(C_{t+1}/C_t)^{-\alpha} \quad (9)$$

A large body of literature in the 1980s tested the pricing Eq. 1 with the SDF given by the consumption model (Eq. 9). See, for example, Hansen and Singleton (1982, 1983), Ferson (1983), and Ferson and Merrick (1987).

More recent work generalizes the consumption-based model to allow for “non-separabilities” in the $U_c(C_t, \cdot)$ function in Eq. 4, as may be implied by the durability of consumer goods, habit persistence in the preferences for consumption, non-separability of preferences across states of nature, and other refinements. Singleton (1990), Ferson (1995), and Cochrane (2001) review this literature; Sarkissian (2003) provides a recent empirical example with references. The rest of this section provides a brief historical overview of empirical work on nonseparable-consumption models.

Dunn and Singleton (1986) and Eichenbaum et al. (1988) developed consumption models with durable goods. Durability introduces nonseparability over time, since the actual consumption at a given date depends on the consumer’s previous expenditures. The consumer optimizes over the current expenditures C_t , accounting for the fact that durable goods purchased today increase consumption at future dates, and thereby lower future marginal utilities. Thus, $U_c(C_t, \cdot)$ in Eq. 4 depends on expenditures prior to date t .

Another form of time nonseparability arises if the utility function exhibits “habit persistence.” Habit persistence means that consumption at two points in time are complements. For example, the utility of current consumption may be evaluated relative to what was consumed in the past, so the previous standard of living influences the utility derived from current consumption. Such models are derived by Ryder and Heal (1973), Becker and Murphy (1988), Sundaresan (1989), Constantinides (1990), and Campbell and Cochrane (1999), among others.

Ferson and Constantinides (1991) model both durability and habit persistence in consumption expenditures. They show that the two combine as opposing effects. In an example based on the utility function of Eq. 8, and where the “memory” is truncated at a single-lag, the derived utility of expenditures is:

$$U(C_t, \cdot) = (1 - \alpha)^{-1} \Sigma_t \beta_t (C_t + bC_{t-1})^{1-\alpha}, \quad (10)$$

where the coefficient b is positive and measures the rate of depreciation if the good is durable and there is no habit persistence. Habit persistence implies that the lagged expenditures enter with a negative effect ($b < 0$). Empirical evidence on similar habit models is provided by Heaton (1993) and Braun et al. (1993), who find evidence for habit in international consumption and returns data.

Consumption expenditure data are highly seasonal, and Ferson and Harvey (1992) argue that the Commerce Department’s X-11 seasonal adjustment program may induce spurious time series behavior in the seasonally adjusted consumption

data that most empirical studies have used. Using data that are not adjusted, they find strong evidence for a seasonal habit model.

Abel (1990) studied a form of habit persistence in which the consumer evaluates current consumption relative to the aggregate consumption in the previous period, and which the consumer takes as exogenous. The idea is that people care about “keeping up with the Joneses.” Campbell and Cochrane (1999) developed another model in which the habit stock is taken as exogenous (or “external”) by the consumer. The habit stock in this case is modeled as a highly persistent weighted average of past aggregate consumptions. This approach results in a simpler and more tractable model, since the consumer’s optimization does not have to take account of the effects of current decisions on the future habit stock. In addition, by modeling the habit stock as an exogenous time series process, Campbell and Cochrane’s model provides more degrees of freedom to match asset market data.

Epstein and Zin (1989, 1991) consider a class of recursive preferences that can be written as: $J_t = F(C_t, CEQ_t(J_{t+1}))$. $CEQ_t(\cdot)$ is a time t “certainty equivalent” for the future lifetime utility J_{t+1} . The function $F(\cdot, CEQ_t(\cdot))$ generalizes the usual expected utility function and may be nontime-separable. They derive a special case of the recursive preference model in which the preferences are:

$$J_t = \left[(1 - \beta)C_t^p + \beta E_t(J_{t+1}^{1-\alpha})^{p/(1-\alpha)} \right]^{1/p}, \quad (11)$$

They show that the IMRS for a representative agent becomes (when $p \neq 0$, $1 - \alpha \neq 0$):

$$m_{t+1} = \left[\beta(C_{t+1}/C_t)^{p-1} \right]^{(1-\alpha)/p} \{R_{m,t+1}\}^{((1-\alpha-p)/p)}. \quad (12)$$

The coefficient of relative risk aversion for timeless consumption gambles is α and the elasticity of substitution for deterministic consumption is $(1 - p)^{-1}$. If $\alpha = 1 - p$, the model reduces to the time-separable power utility model. If $\alpha = 1$, the log utility model of Rubinstein (1976) is obtained. Campbell (1993) shows that the Epstein–Zin model can be transformed to an empirically tractable model without consumption data. He used a linearization of the budget constraint that makes it possible to substitute for consumption in terms of the factors that drive the optimal consumption function. Expected asset returns are then determined by their covariances with the underlying factors.

9.3 Multi-beta Asset Pricing Models

Beta pricing models are a class of asset pricing models that imply the expected returns of securities are related to their sensitivity to changes in the underlying factors that measure the state of the economy. Sensitivity is measured by the securities’ “beta” coefficients. For each of the relevant state variables, there is a

market-wide price of beta measured in the form of an increment to the expected return (a “risk premium”) per unit of beta.

The CAPM represented in Eq. 7 is the premier example of a single-beta pricing model. Multiple-beta models were developed in continuous time by Merton (1973), Breeden (1979), and Cox et al. (1985). Long (1974), Sharpe (1977), Cragg and Malkiel (1982), and Connor (1984). Dybvig (1983), Grinblatt and Titman (1983), and Shanken (1987) provide multi-beta interpretations of equilibrium models in discrete time. Multiple-beta models follow when m_{t+1} can be written as a function of several factors. Equation 3 suggests that likely candidates for the factors are variables that proxy for consumer wealth, consumption expenditures, or the state variables – the sufficient statistics for the marginal utility of future wealth in an optimal consumption–investment plan. A multi-beta model asserts that the expected return is a linear function of several betas, i.e.,

$$E(R_{it+1}) = \delta_0 + \sum_{j=1, \dots, K} \beta_{ij} \delta_j, \quad (13)$$

where the β_{ij} , $j = 1, \dots, K$, are the multiple regression coefficients of the return of asset i on K economy-wide risk factors, f_j , $j = 1, \dots, K$. The coefficient δ_0 is the expected return on an asset that has $\beta_{0j} = 0$, for $j = 1, \dots, K$, i.e., it is the expected return on a zero-(multiple) beta asset. If there is a risk-free asset, then δ_0 is the return for this asset. The coefficient δ_k , corresponding to the k 'th factor has the following interpretation: it is the expected return differential, or premium, for a portfolio that has $\beta_{ik} = 1$ and $\beta_{ij} = 0$ for all $j \neq k$, measured in excess of the zero-beta asset's expected return. In other words, it is the expected return premium per unit of beta risk for the risk factor, k .

A multi-beta model, under certain assumptions, is equivalent to the SDF representation of Eq. 2. This equivalence was first discussed, for the case of the CAPM, by Dybvig and Ingersoll (1982). The general multifactor case is derived by Ferson (1995) and Ferson and Jagannathan (1996), who show that the multi-beta expected return model of Eq. 13 is equivalent to Eq. 2, when the SDF is linear in the factors: $m_{t+1} = a_t + \sum_j b_{jt} f_{jt+1}$.

The logic of the equivalence between multi-beta pricing and the SDF representation of asset pricing models is easily seen using a regression example. Consider a regression of asset returns onto the factors, f_j of the multi-beta model. The regression model is $R_{it+1} = a_i + \sum_j \beta_{ijt} f_{jt} + u_{it+1}$. Substitute the regression equation into the right hand side of Eq. 6 and assume that $\text{Cov}_t(u_{i,t+1}, m_{t+1}) = 0$. The result is:

$$E_t(R_{it+1}) = \delta_{0t} + \sum_{j=1, \dots, K} \beta_{ijt} \left[\text{Cov}_t \left\{ f_{jt+1}, -m_{t+1} \right\} / E_t(m_{t+1}) \right], \quad (14)$$

which is a version of the multi-beta Eq. 13. The market-wide risk premium for factor j is $\delta_{jt} = [\text{Cov}_t \{ f_{jt+1}, -m_{t+1} \} / E_t(m_{t+1})]$. In the special case where the factor f_{jt+1} is a traded asset return, Eq. 14 implies that $\delta_{jt} = E_t(f_{j,t+1}) - \delta_{0t}$; the expected risk premium equals the factor portfolio's expected excess return.

Equation 14 is useful because it provides intuition about the signs and magnitudes of expected risk premiums for particular factors. The intuition is essentially the same as in Eq. 6. If a risk factor f_{jt+1} is negatively correlated with m_{t+1} , the model implies that a positive risk premium is associated with that factor beta. A factor that is negatively related to marginal utility should carry a positive premium, because the big payoffs disappointingly come when the value of payoffs is low. This implies a low present value, and thus a high expected return. With a positive covariance the opposite occurs. If the factor is high when payoffs are highly valued, assets with a positive beta on the factor have a payoff distribution that is “better” than risk free. Thus, the expected return premium is negative, and such assets can have expected returns below that of a risk-free asset.

9.4 Relation to Mean–Variance Efficiency

The concept of a “minimum-variance portfolio” is central in the asset pricing literature. A portfolio R_{pt+1} is minimum variance if and only if no portfolio with the same expected return has a smaller variance. Roll (1977) and others have shown that a portfolio is minimum variance if and only if a single-beta pricing model holds, using the portfolio as the risk factor.¹ According to the CAPM, the market portfolio with return R_{mt+1} is minimum variance. If investors are risk averse, the CAPM also implies that R_{mt+1} is on the positively sloped portion of the minimum-variance frontier, or “mean–variance efficient.” This implies that the coefficient δ_1 in Eq. 7 is positive, which says that there is a positive tradeoff between market risk and expected return when investors are risk averse.

Multiple-beta asset pricing models imply that combinations of particular portfolios are minimum-variance efficient. Equation 13 is equivalent to the statement that a combination of K factor-portfolios is minimum-variance efficient, when the factors are traded assets. This result is proved by Grinblatt and Titman (1987), Shanken (1987), and Huberman et al. (1987). The correspondence between multi-beta pricing and mean variance efficiency is exploited by Jobson and Korkie (1982), Gibbons et al. (1989), Kandel and Stambaugh (1989), and Ferson and Siegel (2005), among others, to develop tests of multi-beta models based on mean variance efficiency.

9.5 Factor Models

A beta pricing model has no empirical content until the factors are specified, since there will almost always be a minimum-variance portfolio which satisfies Eq. 13, with $K = 1$. Therefore, the empirical content of the model is the discipline imposed

¹It is assumed that the portfolio R_{pt+1} is not the global minimum-variance portfolio; that is, the minimum variance over all levels of expected return. This is because the betas of all assets on the global minimum-variance portfolio are identical.

in selecting the factors. There have been four main approaches to finding empirical factors. The first approach is to specify empirical proxies for factors specified by the theory. For example, the CAPM says that the “market portfolio” of all capital assets is the factor, and early studies concentrated on finding good measures for the market portfolio. A second approach is to use factor analytic or principal components methods. This approach is motivated by the APT, as described below. A third approach chooses the risk factors as economic variables or portfolios, based on intuition such as that provided by Eqs. 3 and 4. With this approach, likely candidates for the factors are proxies for consumer wealth, consumer expenditures, and variables that may be sufficient statistics for the marginal utility of future wealth in an optimal consumption–investment plan. For examples of this approach, see Chen et al. (1986), Ferson and Harvey (1991), Campbell (1993), and Cochrane (1996). A fourth approach to factor selection forms portfolios by ranking stocks on firm characteristics that are correlated with the cross-section of average returns. For example, Fama and French (1993, 1996) use the ratio of book value to market price, and the relative market value (size) of the firm to form their “factors.”

Lo and MacKinlay (1990), MacKinlay (1995), and Ferson et al. (1999) provide critiques of the approach of sorting stocks on empirically motivated characteristics in order to form asset pricing factors. Lo and MacKinlay examine the approach as a version of data mining. MacKinlay argues that the factors generated in this fashion by Fama and French (1993, 1996) are statistically unlikely to reflect market risk premiums. Ferson, Sarkissian, and Simin show that a hypothetical characteristic, bearing an anomalous relation to returns, but completely unrelated to risk, can be repackaged as a spurious “risk factor” with this approach. Berk (1995) emphasizes that the price of a stock is the value of its future cash flows discounted by future returns, so an anomalous pattern in the cross-section of returns would produce a corresponding pattern in ratios of cash flow to price. Some of the most empirically powerful characteristics for the cross-sectional prediction of stock returns are ratios, with market price per share in the denominator. However, patterns that are related to the cross-section of asset risks are also likely to be captured by sorting stocks on such ratios. Thus, the approach of sorting stocks on patterns in average returns to form factors is potentially dangerous, because it is likely to “work” when it “should” work, and it is also likely to work when it should not. At the time this chapter was written the controversy over such empirically motivated factors was unresolved.

9.6 Factor Models and the Arbitrage Pricing Model

The Arbitrage Pricing Model based on the APT of Ross (1976) is an example of a multiple-beta asset pricing model, although in the APT Eq. 13 is an approximation. The expected returns are approximately a linear function of the relevant betas as the number of securities in the market grows without bound. Connor (1984) provided sufficient conditions for Eq. 13 to hold exactly in an economy with an infinite number of assets, in general equilibrium. This version of the multiple-beta model, the exact APT, has received wide attention in the finance literature. See Connor and

Korajczyk (1988), Lehmann and Modest (1988), Chen (1983), and Burmeister and McElroy (1988) for discussions on estimating and testing the model when the factor realizations are not observable, under auxiliary assumptions.

This section describes the Arbitrage Pricing Theory (APT) of Ross (1976), and how it is related to factor models and to the general SDF representation for asset pricing models, as in Eq. 2. For this purpose, we suppress the time subscripts and related notation. Assume that the following data-generating model describes equity returns in excess of a risk-free asset:

$$r_i = E(r_i) + \beta'_{if} + e_i \quad (15)$$

where $E(f) = 0 = E(e_{if})$, all i , and $f_t = F_t - E(F_t)$ are the unexpected factor returns. We can normalize the factors to have the identity as their covariance matrix; the β_i absorb the normalization. The $N \times N$ covariance matrix of the asset returns can then be expressed as:

$$\text{Cov}(R) \equiv \Sigma = BB' + V, \quad (16)$$

where V is the covariance matrix of the residual vector, $\mathbf{e}$, B is the $N \times K$ matrix of the vectors, β_i , and Σ is assumed to be nonsingular for all N . An “exact” factor structure assumes that V is diagonal. An approximate factor model, as described by Chamberlain (1983) and Chamberlain and Rothschild (1983), assumes that the eigenvalues of V are bounded as $N \rightarrow \infty$, while the K nonzero-eigenvalues of BB' become infinite as $N \rightarrow \infty$. Thus, the covariance matrix Σ has K unbounded and $N-K$ bounded eigenvalues, as N becomes large.

The factor model represented in Eq. 16 decomposes the variances of returns into “pervasive” and “nonsystematic” risks. If $\mathbf{x}$ is an N -vector of portfolio weights, the portfolio variance is $\mathbf{x}'\Sigma\mathbf{x}$, where $\lambda_{\max}(\Sigma)\mathbf{x}'\mathbf{x} \geq \mathbf{x}'\Sigma\mathbf{x} \geq \lambda_{\min}(\Sigma)\mathbf{x}'\mathbf{x}$, $\lambda_{\min}(\Sigma)$ being the smallest eigenvalue of Σ and $\lambda_{\max}(\Sigma)$ being the largest. Following Chamberlain (1983), a portfolio is “well diversified” if $\mathbf{x}'\mathbf{x} \rightarrow 0$ as N grows without bound. For example, an equally weighted portfolio is well diversified; in this case $\mathbf{x}'\mathbf{x} = (1/N) \rightarrow 0$. The bounded eigenvalues imply that V captures the component of portfolio risk that is not pervasive or systematic, in the sense that this part of the variance vanishes in a well-diversified portfolio. The exploding eigenvalues of BB' imply that the common factor risks are pervasive, in the sense that they remain in a large, well-diversified portfolio.

The arbitrage pricing theory of Ross (1976) asserts that $\alpha'\alpha < \infty$ as N grows without bound, where α is the N vector of “alphas,” or expected abnormal returns, measured as the differences between the left and right hand sides of Eq. 13, using the APT factors in the multi-beta model. The alphas are the differences between the assets’ expected returns and the returns predicted by the multi-beta model, also called the “pricing errors.” The Ross APT implies that the multi-beta model’s pricing errors are “small” on average, in a large market. If $\alpha'\alpha < \infty$ as N grows, then the cross-asset average of the squared pricing errors, $(\alpha'\alpha)/N$ must go to 0 as N grows.

The pricing errors in a beta pricing model are related to those of a SDF representation. If we define $\alpha_m = E(m\mathbf{R} - 1)$, where m is linear in the APT factors, then it

follows that $\alpha_m = E(m)\alpha$; the beta pricing and stochastic discount factor alphas are proportional, where the risk-free rate determines the constant of proportionality. Provided that the risk-free rate is bounded above 100 %, then $E(m)$ is bounded, and $\alpha'\alpha$ is bounded above if and only if $\alpha'_m\alpha_m$ is bounded above. Thus, the Ross APT has the same implications for the pricing errors in the SDF and beta pricing paradigms.

The “exact” version of the APT derived by Connor (1984) asserts that $\alpha'\alpha \rightarrow 0$ as N grows without bound, and thus the pricing errors of all assets go to zero as the market gets large. Chamberlain (1983) shows that the exact APT is equivalent to the statement that all minimum-variance portfolios are well diversified, and are thus combinations of the APT factors. In this case, we have $E(m\mathbf{R} - 1) = 0$ when m is linear in the APT factors, and a combination of the factors is a minimum-variance efficient portfolio in the large market.

9.7 Summary

The asset pricing models of financial economics are based on an assumption that rules out arbitrage opportunities, or they rely on explicit equilibrium conditions. Empirically, there are three central representations. The first is the minimum-variance efficiency of a portfolio. The second is the beta pricing model stated in terms of risk factors, and the third is the SDF representation. These three representations are closely related, and become equivalent under ancillary assumptions. Together they provide a rich and flexible framework for empirical analysis.

References

- Abel, A.B. 1990. Asset prices under habit formation and catching up with the Joneses. *American Economic Review* 80: 38–42.
- Barberis, N., and A. Shleifer. 2003. Behavioral finance. In *Handbook of the economics of finance*, ed. G.M. Constantinides, M. Harris, and R.M. Stulz. North Holland: Elsevier.
- Becker, G., and K.M. Murphy. 1988. A theory of rational addiction. *Journal of Political Economy* 96: 675–700.
- Beja, A. 1971. The structure of the cost of capital under uncertainty. *Review of Economic Studies* 4: 359–369.
- Berk, J. 1995. A critique of size-related anomalies. *Review of Financial Studies* 8: 275–286.
- Bhattacharya, S. 1981. Notes on multiperiod valuation and the pricing of options. *Journal of Finance* 36: 163–180.
- Braun, P., W.E. Ferson, and G.M. Constantinides. 1993. Time nonseparability in aggregate consumption: International evidence. *European Economic Review* 37: 897–920.
- Breeden, D. 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics* 7: 265–296.
- Brennan, M.J., and A. Kraus. 1978. Necessary conditions for aggregation in securities markets. *Journal of Financial and Quantitative Analysis* 407–418.
- Burmeister, E., and M.B. McElroy. 1988. Joint estimation of factor sensitivities and risk premia for the Arbitrage Pricing Theory. *Journal of Finance* 43 (3): 721–733.

- Campbell, J.Y. 1993. Intertemporal asset pricing without consumption data. *American Economic Review* 83: 487–512.
- Campbell, J.Y., and J.H. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107: 205–251.
- Chamberlain, G. 1983. Funds, factors and diversification in arbitrage pricing models. *Econometrica* 51: 1305–1324.
- Chamberlain, G., and M. Rothschild. 1983. Arbitrage, factor structure and mean variance analysis on large asset markets. *Econometrica* 51: 1281–1304.
- Chen, N.F. 1983. Some empirical tests of the theory of arbitrage pricing. *Journal of Finance* 38: 1393–1414.
- Chen, N., R. Roll, and S. Ross. 1986. Economic forces and the stock market. *Journal of Business* 59: 383–403.
- Cochrane, J.H. 1996. A cross-sectional test of a production based asset pricing model, working paper. *Journal of Political Economy* 104: 572–621.
- . 2001. *Asset pricing*. Princeton: Princeton University.
- Connor, G. 1984. A unified beta pricing theory. *Journal of Economic Theory* 34: 13–31.
- Connor, G., and R. Korajczyk. 1988. Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics* 21: 255–290.
- Constantinides, G.M. 1982. Intertemporal asset pricing with heterogeneous consumers and without demand aggregation. *Journal of Business* 55: 253–267.
- . 1990. Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy* 98: 519–543.
- Constantinides, G.M., and D. Duffie. 1996. Asset pricing with heterogeneous consumers. *Journal of Political Economy* 104: 219–240.
- Cox, J.C., J.E. Ingersoll, and S.A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53: 385–408.
- Cragg, J.G., and B.G. Malkiel. 1982. *Expectations and the structure of share prices*. Chicago: University of Chicago Press.
- Dunn, K.B., and K.J. Singleton. 1986. Modelling the term structure of interest rates under non-separable utility and durability of goods. *Journal of Financial Economics* 17: 27–55.
- Dybvig, P.H. 1983. An explicit bound on individual assets' deviations from APT pricing in a finite economy. *Journal of Financial Economics* 12: 483–496.
- Dybvig, P., and J. Ingersoll. 1982. Mean variance theory in complete markets. *Journal of Business* 55: 233–252.
- Eichenbaum, M.S., L.P. Hansen, and K.J. Singleton. 1988. A time series analysis of representative agent models of consumption and leisure choices under uncertainty. *Quarterly Journal of Economics* 103: 51–78.
- Epstein, L.G., and S.E. Zin. 1989. Substitution, risk aversion and the temporal behavior of asset returns: A theoretical approach. *Econometrica* 57: 937–970.
- . 1991. Substitution, risk aversion and the temporal behavior of asset returns. *Journal of Political Economy* 99: 263–286.
- Fama, E.F., and K.R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56.
- . 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance* 51: 55–87.
- Ferson, W.E. 1983. Expectations of real interest rates and aggregate consumption: Empirical tests. *Journal of Financial and Quantitative Analysis* 18: 477–497.
- . 1995. Theory and empirical testing of asset pricing models. In *Handbooks in operations research and management science*, ed. J. Maksimovic and Ziemba. Amsterdam: Elsevier.
- Ferson, W.E., and G.M. Constantinides. 1991. Habit persistence and durability in aggregate consumption: Empirical tests. *Journal of Financial Economics* 29: 199–240.
- Ferson, W.E., and C.R. Harvey. 1991. The variation of economic risk premiums. *Journal of Political Economy* 99: 385–415.
- . 1992. Seasonality and consumption based asset pricing models. *Journal of Finance* 47: 511–552.

- Ferson, W.E., and R. Jagannathan. 1996. Econometric evaluation of asset pricing models. In *Handbook of statistics: Volume 14: Statistical methods in finance*, ed. G.S. Maddala and C.R. Rao. Amsterdam.
- Ferson, W.E., and J.J. Merrick. 1987. Non-stationarity and stage of the business cycle effects in consumption-based asset pricing relations. *Journal of Financial Economics* 18: 127–146.
- Ferson, W.E., and Siegel, A.F. (2005). Testing portfolio efficiency with conditioning information. Working paper, Boston College.
- Ferson, W.E., S. Sarkissian, and T. Simin. 1999. The alpha factor asset pricing model: A parable. *Journal of Financial Markets* 2: 49–68.
- Gibbons, M.R., S.A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* 57: 1121–1152.
- Gorman, W.M. 1953. Community preference fields. *Econometrica* 21: 63–80.
- Grinblatt, M., and S. Titman. 1983. Factor pricing in a finite economy. *Journal of Financial Economics* 12: 497–508.
- . 1987. The relation between mean-variance efficiency and arbitrage pricing. *Journal of Business* 60: 97–112.
- Grossman, S., and R.J. Shiller. 1982. Consumption correlatedness and risk measurement in economies with nontraded assets and heterogeneous information. *Journal of Financial Economics* 10: 195–210.
- Hansen, L.P., and S.F. Richard. 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* 55: 587–613.
- Hansen, L.P., and K.J. Singleton. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50: 1269–1285.
- . 1983. Stochastic consumption, risk aversion and the temporal behavior of asset returns. *Journal of Political Economy* 91: 249–266.
- Harrison, M., and D. Kreps. 1979. Martingales and arbitrage in multi-period securities markets. *Journal of Economic Theory* 20: 381–408.
- Heaton, J. 1993. The interaction between time-non-separable preferences and time aggregation. *Econometrica* 61 (2): 353–385.
- Huberman, G., S.A. Kandel, and R.F. Stambaugh. 1987. Mimicking portfolios and exact arbitrage pricing. *Journal of Finance* 42: 1–10.
- Jobson, J.D., and R. Korkie. 1982. Potential performance and tests of portfolio efficiency. *Journal of Financial Economics* 10: 433–466.
- Kandel, S., and R.F. Stambaugh. 1989. A mean variance framework for tests of asset pricing models. *Review of Financial Studies* 2: 125–156.
- Lee, C.F., C. Wu, and K.C.J. Wei. 1990. The heterogeneous investment horizon and the capital asset pricing model: Theory and implications. *Journal of Financial and Quantitative Analysis* 25: 361–376.
- Lehmann, B.N., and D.M. Modest. 1988. The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics* 21: 213–254.
- Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–37.
- Lo, A.W., and A.C. MacKinlay. 1990. Data snooping in tests of financial asset pricing models. *Review of Financial Studies* 3: 431–467.
- Long, J. 1974. Stock prices, inflation, and the term structure of interest rates. *Journal of Financial Economics* 1: 131–170.
- Lucas, R.E., Jr. 1978. Asset prices in an exchange economy. *Econometrica* 46: 1429–1445.
- MacKinlay, A.C. 1995. Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics* 38: 3–28.
- Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867–887.
- Mossin, J. 1966. Equilibrium in a capital asset market. *Econometrica* 34 (4): 768–783.
- Roll, R. 1977. A critique of the asset pricing theory's tests-part 1: On past and potential testability of the theory. *Journal of Financial Economics* 4: 129–176.

- Ross, S.A. 1976. The arbitrage pricing theory of capital asset pricing. *Journal of Economic Theory* 13: 341–360.
- Ross, S. 1977. Risk, return and arbitrage. In *Risk and return in finance*, ed. I. Friend and J. Bicksler. Cambridge, MA: Ballinger.
- Rubinstein, M. 1974. An aggregation theorem for securities markets. *Journal of Financial Economics* 1: 225–244.
- . 1976. The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics and Management Science* 7: 407–425.
- Ryder, H.E., Jr., and G.M. Heal. 1973. Optimal growth with intertemporally dependent preferences. *Review of Economic Studies* 40: 1–31.
- Sarkissian, S. 2003. Incomplete consumption risk sharing and currency premiums. *Review of Financial Studies* 16: 983–1005.
- Shanken, J. 1987. Multivariate proxies and asset pricing relations: Living with the roll critique. *Journal of Financial Economics* 18: 91–110.
- Sharpe, W.F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- . 1977. The capital asset pricing model: A multi-beta interpretation. In *Financial decision making under uncertainty*, ed. H. Levy and M. Sarnat. New York: Academic.
- Singleton, K.J. 1990. Specification and estimation of intertemporal asset pricing models. In *Handbook of monetary economics*, ed. E. Friedman and F. Hahn. North Holland: Elsevier.
- Sundaresan, S.M. 1989. Intertemporally dependent preferences and the volatility of consumption and wealth. *Review of Financial Studies* 2: 73–89.
- Wilson, R.B. 1968. The theory of syndicates. *Econometrica* 36: 119–131.

Conditional Asset Pricing

10

Wayne E. Ferson

Abstract

Conditional asset pricing studies predictability in the returns of financial assets, and the ability of asset pricing models to explain this predictability. The relation between predictability and asset pricing models is explained and the empirical evidence for predictability is summarized. Empirical tests of conditional asset pricing models are then briefly reviewed.

Keywords

Capital asset pricing models · Conditional expectations · Discount rates · Financial asset returns · Latent variables · Market price of risk · Mean variance efficiency · Minimum-variance portfolios · Multiple-beta models · Predictability · Rational expectations · Stochastic discount factors · Stock prices

10.1 Introduction

Conditional Asset Pricing refers to a subset of Asset Pricing research in financial economics (see ► [Chap. 9, “Asset Pricing Models”](#)). Conditional Asset Pricing focuses on predictability over time in rates of return on financial assets, and the ability of asset pricing models to explain this predictability.

Most asset pricing models are special cases of the fundamental equation:

$$P_t = E_t\{m_{t+1}(P_{t+1} + D_{t+1})\} \quad (1)$$

where P_t is the price of the asset at time t , and D_{t+1} is the amount of any dividends, interest or other payments received at time $t + 1$. The market-wide random variable

W. E. Ferson (✉)
University of Southern California, Los Angeles, CA, USA
e-mail: ferson@marshall.usc.edu

m_{t+1} is the “stochastic discount factor” (SDF). By recursive substitution in Eq. 1, the future price may be eliminated to express the current price as a function of the future cash flows and SDFs only: $P_t = E_t\{\sum_{j=0}^{\infty} \prod_{k=1, \dots, j} m_{t+k} D_{t+j}\}$. Prices are obtained by “discounting” the payoffs, or multiplying by SDFs, so that the expected “present value” of the payoff is equal to the price. A SDF “prices” the assets if Eq. 1 is satisfied, and any particular asset pricing model may be viewed as a specification for the stochastic discount factor.

The notation $E_t\{\cdot\}$ in Eq. 1 denotes the conditional expectation, given a market-wide information set, Ω_t . Empiricists don’t get to see Ω_t , so it is convenient to consider expectations conditional on an observable subset of instruments, Z_t . These expectations are denoted as $E(\cdot|Z_t)$. When Z_t is the null information set, we have the unconditional expectation, denoted as $E(\cdot)$.

Empirical work on conditional asset pricing models typically relies on “rational expectations,” which is the assumption that the expectation terms in the model are mathematical conditional expectations. This carries two important implications. First, it implies that the “law of iterated expectations” can be invoked. This says that the expectation, given coarser information, of the conditional expectation given finer information, is the conditional expectation given the coarser information. For example, taking the expected value of Eq. 1, rational expectations implies that versions of Eq. 1 must hold for the expectations $E(\cdot|Z_t)$ and $E(\cdot)$. Second, rational expectations implies that the differences between realizations of the random variables and the expectations in the model, should be unrelated to the information that the expectations in the model are conditioned on. This leads to implications for the predictability of asset returns.

Define the gross asset return, $R_{it+1} = (P_{it+1} + D_{it+1})/P_{it}$. The return of the asset i may be predictable. For example, a linear regression over time of R_{it+1} on Z_t may have a nonzero slope coefficient. Eq. 1 implies that the conditional expectation of the product of m_{t+1} and R_{it+1} is the constant, 1.0. Therefore, $1 - m_{t+1}R_{it+1}$ should not be predictably different from 0 using any information available at time t . If there is predictability in a return R_{it+1} using any lagged instruments Z_t , the model implies that the predictability is removed when R_{it+1} is multiplied by the correct m_{t+1} . This is the sense in which conditional asset pricing models are asked to “explain” predictable variation in asset returns.

If a conditional asset pricing model fails to explain predictability as described above, there are two possibilities (Fama 1970, 1991). Either the specification of m_{t+1} in the model is wrong, or the use of rational expectations is unjustified. The first instance motivates research on better conditional asset pricing models. The second possibility motivates research on human departures from rationality, and how these show up in asset market prices. For a review of this relatively new field, “behavioral finance,” see Barberis and Shleifer (2003).

Studies of predictability in stock and long-term bond returns typically report regressions that attempt to predict the future returns using lagged variables. These regressions for shorter horizon (monthly, or annual holding period) returns typically have small R -squares, as the fraction of the variance in long-term asset returns that can be predicted with lagged variables over short horizons is small. The R -squares

are larger for longer-horizon (2- to 5-year) returns, because expected returns are considered to be more persistent than returns themselves. Thus, the variance of the expected return accumulates with longer horizons faster than the variance of the return, and the R -squared increases (Fama and French 1988).

Because stock returns are very volatile, small R -squares can mask economically important variation in the expected return. To illustrate, consider a special case of Eq. 1, the simple Gordon (1962) constant-growth model for a stock price: $P = kE/(R - g)$, where P is the stock price, E is the earnings per share, k is the dividend payout ratio, g is the future growth rate of earnings, and R is the discount rate. The discount rate is the required or expected return of the stock. Stocks are long “duration” assets, so a small change in the expected return can lead to a large fluctuation in the asset value. Consider an example where the price-to-earnings ratio, $P/E = 15$, the payout ratio, $k = 0.6$, and the expected growth rate, $g = 3\%$. The expected return, R , is 7%. Suppose there is a shock to the expected return, *ceteris paribus*. In this example a change of 1% in R leads to approximately a 20% change in the asset value.

Of course, it is unrealistic to hold everything else fixed, but the example suggests that small changes in expected returns can produce large and economically significant changes in asset values. Campbell (1991) generalizes the Gordon model to allow for stochastic changes in growth rates, and estimates that changes in expected returns through time may account for about half of the variance of equity index values. Conditional Asset Pricing models focus on these changes in the required or expected rates of return on financial assets.

10.2 The Conditional Capital Asset Pricing Model

The simplest example of a conditional asset pricing model is a conditional version of the Capital Asset Pricing Model (CAPM) of Sharpe (1964):

$$E(R_{it+1}|Z_t) = \gamma_o(Z_t) + \beta_{imt} \gamma_m(Z_t) \quad (2)$$

where R_{it+1} is the rate of return of asset i between times t and $t + 1$, and β_{imt} is the market beta at time t . The market beta is the conditional covariance of the return with the market portfolio divided by the conditional variance of the market portfolio; that is, the slope coefficient in a conditional regression of the asset return on that of the market, conditional on the information at time t . Z_t is the conditioning information, assumed to be publicly available at time t . The term $\gamma_m(Z_t)$ represents the risk premium for market beta, and $\gamma_o(Z_t)$ is the expected return of all portfolios with market betas equal to zero. If there is a risk-free asset available at time t , then its rate of return equals $\gamma_o(Z_t)$.

Sharpe (1964) did not explicitly put the conditioning information, Z_t , into his derivation of the CAPM. The original development was cast in a single-period partial equilibrium model. However, it is natural to interpret the expectations in the model as reflecting a consensus of well-informed analysts’ opinion – conditional

expectations given their information – and Sharpe’s subsequent writings indicated this intent (e.g. Sharpe 1984). The multiple-beta intertemporal models of Merton (1973) and Cox–Ingersoll–Ross (1985) accommodate conditional expectations explicitly. Merton (1973, 1980) and Cox–Ingersoll–Ross also showed how conditional versions of the CAPM may be derived as special cases of their models.

Roll (1977) and others have shown that a portfolio is “minimum variance” if and only if a model like Eq. 2 fits the expected returns for all the assets i , using the minimum-variance portfolio as R_{mt+1} . A portfolio is minimum variance if and only if no portfolio with the same expected return has a smaller variance. According to the CAPM, the market portfolio with return R_{mt+1} is minimum variance. If investors are risk averse, the CAPM also implies that the market portfolio is “mean-variance efficient,” which says that $\gamma_m(Z_t)$ in Eq. 2 is positive. In the CAPM, risk-averse investors choose portfolios that have the maximum expected return, given the variance. This implies that there is a positive tradeoff between market risk, as measured by β_{imt} , and the expected return on individual assets, when investors are risk averse. In the conditional CAPM, mean–variance efficiency is defined relative to the conditional expectations and conditional variances of returns. Hansen and Richard (1987) and Ferson and Siegel (2001) describe theoretical relations between conditional and “unconditional” versions of mean–variance efficiency.

The conditional CAPM may be expressed in the SDF representation given by Eq. 1 as: $m_{t+1} = c_{0t} - c_{1t}R_{mt+1}$. In this case, the coefficients c_{0t} and c_{1t} are specific measurable functions of the information set Z_t , depending on the first and second conditional moments of the returns. To implement the model empirically, it is necessary to specify functional forms for c_{0t} and c_{1t} . Shanken (1990) suggests approximating the coefficients using linear functions, and this approach is followed by Cochrane (1996), Jagannathan and Wang (1996), and other authors.

10.3 Evidence for Return Predictability

Conditional asset pricing presumes the existence of some return predictability. There should be instruments Z_t for which $E(m_{t+1}|Z_t)$ or $E(R_{t+1}|Z_t)$ vary over time, in order for $E(m_{t+1}R_{t+1} - 1|Z_t) = 0$ to have empirical bite. At one level, this is easy. Since $E(m_{t+1}|Z_t)$ should be the inverse of a risk-free return, all we need for the first condition to bite is observable risk-free rates that vary over time. Indeed, a short-term interest rate is one of the most prominent of the lagged instruments used to represent Z_t in empirical work. Ferson (1989) shows that the behavior of stock returns and short-term interest rates, as documented by Fama and Schwert (1977), imply that conditional covariances of returns with m_{t+1} must also vary over time.

Interest in predicting security returns is probably as old as the security markets themselves. Fama (1970) reviews the early evidence and Schwert (2003) reviews anomalies in asset pricing based on predictability. It is useful to distinguish, following Fama (1970), predictability based on the information in past returns (“weak form”) from predictability based on lagged economic variables that are public information, not limited to past prices and returns (“semi-strong” form).

A large body of literature studies weak-form predictability, focusing on serial dependence in returns. High-frequency serial dependence, such as daily or intra-day patterns, are often considered to represent the effects of market microstructure, such as bid–ask spreads (e.g. Roll 1984) and non-synchronous trading of the stocks in an index (e.g. Scholes and Williams, 1977). Serial dependence may also represent predictable changes in the expected returns.

Conrad and Kaul (1989) report serial dependence in weekly returns. Jegadeesh and Titman (1993) find that relatively high-return recent “winner” stocks tend to repeat their performance over 3- to 9-month horizons. DeBondt and Thaler (1985) find that past high-return stocks perform poorly over the next 5 years, and Fama and French (1988) find negative serial dependence over 2- to 5-year horizons. These serial dependence patterns motivate a large number of studies, which attempt to assess the economic magnitude and statistical robustness of the implied predictability, or to explain the predictability as an economic phenomenon. For a summary of this literature subsequent to Fama (1970), see Campbell et al. (1997). Research in this area continues, and it’s fair to say that the jury is still out on the issue of predictability using lagged returns.

A second body of literature studies semi-strong form predictability using other lagged, publicly available information variables as instruments. Fama and French (1989) assemble a list of variables from studies in the early 1980s, which as of this writing remain the workhorse instruments for conditional asset pricing models. In addition to the level of a short-term interest rate, as mentioned above, the variables include the lagged dividend yield of a stock market index, a yield spread of long-term government bonds relative to short-term bonds, and a yield spread of low-grade (high-default risk and low liquidity) corporate bonds over high-grade corporate bonds. In addition, studies often use the lagged excess return of a medium-term over a short-term Treasury bill (Campbell 1987; Ferson and Harvey 1991). Additional instruments include an aggregate book-to-market ratio (Pontiff and Schall 1998) and lagged consumption-to-wealth ratios (Lettau and Ludvigson 2001a). Of course, many other predictor variables have been proposed and more will doubtless be proposed in the future.

Predictability using lagged instruments remains controversial, and there are some good reasons the measured predictability could be spurious. Studies have identified various statistical biases in predictive regressions (e.g. Hansen and Hodrick 1980; Stambaugh 1999; Ferson et al. 2003), and have questioned the stability of predictive relations across economic regimes (e.g. Kim et al. 1991; or Paye and Timmermann 2003) and raised the possibility that the lagged instruments arise solely through data mining (e.g. Lo and MacKinlay 1990; Foster et al. 1997).

A reasonable response to these concerns is to see if the predictive relations hold out-of-sample. This kind of evidence is also mixed. Some studies find support for predictability in step-ahead or out-of-sample exercises (e.g. Fama and French 1989; Pesaran and Timmermann 1995). Similar instruments show some ability to predict returns outside the United States, where they were originally studied (e.g. Harvey 1991; Solnik 1993; Ferson and Harvey 1993, 1999). However, other studies conclude that predictability using the standard lagged instruments does not hold in more

recent samples (e.g. Goyal and Welch 2003; Simin 2002). It seems that research on the predictability of security returns will always be interesting, and conditional asset pricing models should be useful in framing many future investigations of these issues.

10.4 Tests of Conditional CAPMs

Empirical studies have rejected versions of the CAPM that ignore lagged variables. This evidence, and mounting evidence of predictable variation in the distribution of security returns led to empirical work on conditional versions of the CAPM starting in the early 1980s. An example from Eq. 2 illustrates the implications of the conditional CAPM for predictability in returns. Rational expectations implies that the actual return differs from the conditional expected value by an error term, u_{it+1} , which is orthogonal to the information at time t . If the actual returns are predictable using information in Z_t , the model implies that either the betas or the premiums ($\gamma_m(Z_t)$ and $\gamma_o(Z_t)$), are changing as functions of Z_t , and the time variation in those functions should track the predictable components of asset returns. If the time variation in $\gamma_m(Z_t)$ and $\gamma_o(Z_t)$ can be modeled, the conditional CAPM can be tested by examining its ability to explain the predictability in returns.

The earliest empirical tests along these lines were the “latent variable models,” developed by Hansen and Hodrick (1983) and Gibbons and Ferson (1985), and later refined by Campbell (1987) and Ferson et al. (1993). These models allow time varying expected returns, but maintain the assumption that the conditional betas are fixed parameters over time.

Consider the conditional representation of the CAPM. Let $r_{it+1} = R_{it+1} - R_{0t+1}$, and similarly for the market return, where R_{0t+1} is the gross, zero beta return. The conditional CAPM may then be stated for the vector of excess returns r_{t+1} , as $E(r_{t+1}|Z_t) = \beta E(r_{mt+1}|Z_t)$, where β is the vector of assets’ betas. Let r_{1t} be any reference asset excess return with nonzero beta, β_1 , so that $E(r_{1t+1}|Z_t) = \beta_1 E(r_{mt+1}|Z_t)$. Solving this expression for $E(r_{mt+1}|Z_t)$ and substituting, we have $E(r_{t+1}|Z_t) = \mathbf{C}E(r_{1t+1}|Z_t)$, where $\mathbf{C} = (\beta/\beta_1)$. and $/$ denotes element-by-element division. The expected market risk premium is now a latent variable in the model, and $\mathbf{C}$ is the N -vector of the model parameters. Gibbons and Ferson (1985) argued that the latent variable model is attractive in view of the difficulties associated with measuring the true market portfolio of the CAPM, but Wheatley (1989) emphasized that it remains necessary to assume that ratios of the betas measured with respect to the unobserved market portfolio, are constant parameters.

Campbell (1987) and Ferson and Foerster (1994) show that a single-beta latent variable model is rejected by the data. This rejects the hypothesis that there is a conditional minimum-variance portfolio such that the ratios of conditional betas on this portfolio are fixed parameters. Therefore, the empirical evidence suggests that conditional asset pricing models should be consistent with either (1) a time varying beta, or (2) more than one beta for each asset.

Conditional CAPMs with time varying betas are examined by Harvey (1989), replacing the constant beta assumption with the assumption that the ratio of the expected market premium to the conditional market variance is a fixed parameter: $E(r_{mt+1}|Z_t)/\text{Var}(r_{mt+1}|Z_t) = \gamma$. Then, the conditional expected returns may be written according to the conditional CAPM as $E(r_{t+1}|Z_t) = \gamma \text{Cov}(r_{t+1}, r_{mt+1}|Z_t)$. Harvey's version of the conditional CAPM is motivated from Merton's (1980) model, in which the ratio γ , called the "market price of risk," is equal to the relative risk aversion of a representative investor in equilibrium. Harvey also assumes that the conditional expected risk premium on the market (and the conditional market variance, given fixed γ) is a linear function of the instruments: $E(r_{mt+1}|Z_t) = \delta_m' Z_t$, where δ_m is a coefficient vector. He rejects this version of the conditional CAPM for monthly data in the United States. In Harvey (1991), the same formulation is rejected when applied to a world market portfolio and monthly data on the stock markets of 21 developed countries.

Lettau and Ludvigson (2001b) examine a conditional CAPM with time varying betas and risk premiums, using rolling time-series and cross-sectional regression methods. They condition the model on a lagged, consumption-to-wealth ratio, and find that the conditional CAPM works better for explaining the cross-section of monthly stock returns.

10.5 Multi-beta Conditional Asset Pricing Models

A multi-beta asset pricing model essentially expands Eq. 2 to allow for multiple sources of risk and expected return. Such a model asserts that the expected return is a linear function of several betas, i.e.

$$E_t(R_{it+1}) = \lambda_{0t} + \sum_{j=1, \dots, K} \beta_{ijt} \lambda_{jt}, \quad (3)$$

where the β_{ijt} , $j = 1, \dots, K$, are the conditional multiple regression coefficients of the return of asset i on K risk factors, f_{jt+1} , $j = 1, \dots, K$. The coefficient λ_{0t} is the expected return on an asset that has $\beta_{0jt} = 0$, for $j = 1, \dots, K$; i.e. it is the expected return on a zero-(multiple-) beta asset. If there is a risk-free asset, then λ_{0t} is the return of this asset. The coefficient λ_{kt} , corresponding to the k 'th factor has the following interpretation: it is the expected return differential, or premium, for a portfolio that has $\beta_{ikt} = 1$ and $\beta_{ijt} = 0$ for all $j \neq k$, measured in excess of the zero-beta asset's expected return. In other words, it is the expected return premium per unit of beta risk for the risk factor, k . Multiple-beta models follow when m_{t+1} can be written as a conditional linear function of the K factors, as shown by Ferson and Jagannathan (1996).

Bansal and Viswanathan (1993) developed conditional versions of the CAPM and multiple-factor models in which the stochastic discount factor m_{t+1} is a nonlinear function of the market or factor returns. Using nonparametric methods, they find evidence to support the nonlinear versions of the models. Bansal et al. (1993) compare the performance of nonlinear models with linear models, using data on

international stocks, bonds, and currency returns, and they find that the nonlinear models perform better. Farnsworth et al. (2002) compared the empirical performance of a large set of conditional asset pricing models using the SDF representation.

Conditional multiple-beta models with constant betas are examined empirically by Ferson and Harvey (1991), Evans (1994), and Ferson and Korajczyk (1995), who find that while such models are rejected using the usual statistical tests, they still capture a large fraction of the predictability of stock and bond returns over time. Allowing for time varying betas, these studies find that the time variation in betas contributes a relatively small amount to the time variation in expected asset returns, while time variation in the risk premium are relatively more important.

While time variation in conditional betas is not as important as time variation in expected risk premiums, from the perspective of modeling predictable time variation in asset returns, this does not imply that beta variation is empirically unimportant. From the perspective of modeling the cross-sectional variation in “unconditional” expected asset returns, beta variation over time may be empirically very important. This idea was first explored by Chan and Chen (1988). To see how this works, consider the unconditional expected excess return vector, obtained from the model as $E\{E(\mathbf{r}|Z)\} = E\{\lambda(Z)\boldsymbol{\beta}(Z)\} = E(\lambda)E(\boldsymbol{\beta}) + \text{Cov}(\lambda(Z), \boldsymbol{\beta}(Z))$. Viewed as a cross-sectional relation, the term $\text{Cov}(\lambda(Z), \boldsymbol{\beta}(Z))$ may differ significantly in a cross-section of assets. Therefore the implications of a conditional version of the CAPM for the cross-section of unconditional expected returns may depend importantly on common time variation in betas and expected market risk premiums.

Jagannathan and Wang (1996) used the conditional CAPM to derive a particular “unconditional” 2-factor model. They show that $m_{t+1} = \alpha_0 + \alpha_1 E(r_{mt+1}|\Omega_t) + R_{mt+1}$, where Ω_t denotes the information set of investors and α_0 and α_1 are fixed parameters, is a valid SDF in the sense that $E(R_{t,t+1}m_{t+1}) = 1$ for this choice of m_{t+1} . Assuming that $E(r_{mt+1}|Z_t)$ is a linear function of Z_t , they find that their version of the model explains the cross-section of unconditional expected returns better than an unconditional version of the CAPM.

New empirical tests of the conditional CAPM and multiple-beta models, using the multi-beta representation and SDF representations, continue to appear regularly in the literature. Future studies will continue to refine the relationships among the various empirical specifications. Research on the predictability of security returns will always be interesting, and Conditional Asset Pricing Models should be useful in framing many future investigations of these issues.

References

- Bansal, R., and S. Viswanathan. 1993. No arbitrage and arbitrage pricing: a new approach. *Journal of Finance* 48: 1231–1262.
- Bansal, R., D. Hsieh, and S. Viswanathan. 1993. A new approach to international arbitrage pricing. *Journal of Finance* 48: 1719–1747.
- Barberis, N., and A. Shleifer. 2003. A survey of behavioral finance. In *Handbook of the Economics of Finance*, ed. R. Stulz. Elsevier: North Holland.
- Campbell, J.Y. 1987. Stock returns and the term structure. *Journal of Financial Economics* 18: 373–399.

- . 1991. A variance decomposition for stock returns. *Economic Journal* 101: 157–179.
- Campbell, J.Y., A. Lo, and A.C. MacKinlay. 1997. *The econometrics of financial markets*. Princeton: Princeton University Press.
- Chan, K.C., and N.F. Chen. 1988. An unconditional test of asset pricing and the role of firm size as an instrumental variable for risk. *Journal of Finance* 63: 431–468.
- Cochrane, J.H. 1996. A cross-sectional test of a production based asset pricing model. *Journal of Political Economy* 104 (3): 572–621.
- Conrad, J., and G. Kaul. 1989. Mean reversion in short-horizon expected returns. *Review of Financial Studies* 2: 225–240.
- Cox, J.C., J.E. Ingersoll, and S.A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53: 385–408.
- DeBondt, W., and R. Thaler. 1985. Does the stock market overreact? *Journal of Finance* 40: 793–805.
- Evans, M.D.D. 1994. Expected returns, time-varying risk and risk premia. *Journal of Finance* 49 (2): 655–679.
- Fama, E.F. 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance* 25: 383–417.
- . 1991. Efficient capital markets II. *Journal of Finance* 46: 1575–1618.
- Fama, E., and K. French. 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96: 246–273.
- . 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25: 23–49.
- Fama, E.F., and G.W. Schwert. 1977. Asset returns and inflation. *Journal of Financial Economics* 5: 115–146.
- Farnsworth, H., W.E. Ferson, D. Jackson, and S. Todd. 2002. Performance evaluation with stochastic discount factors. *Journal of Business* 75: 473–504.
- Ferson, W.E. 1989. Changes in expected security returns, risk and the level of interest rates. *Journal of Finance* 44: 1191–1217.
- Ferson, W.E., and S.R. Foerster. 1994. Finite sample properties of the generalized methods of moments tests of conditional asset pricing models. *Journal of Financial Economics* 36: 29–56.
- Ferson, W.E., and C.R. Harvey. 1991. The variation of economic risk premiums. *Journal of Political Economy* 99: 385–415.
- . 1993. The risk and predictability of international equity returns. *Review of Financial Studies* 6: 527–566.
- . 1999. Economic, financial and fundamental global risk in and out of EMU. *Swedish Economic Policy Review* 6: 123–184.
- Ferson, W. E., & R. Jagannathan. (1996). Econometric evaluation of asset pricing models. In: G.S. Maddala and C.R. Rao (eds.) *Handbook of statistics*. 14 Chapter 1, 1–30, Amsterdam, the Netherlands.
- Ferson, W.E., and R.A. Korajczyk. 1995. Do arbitrage pricing models explain the predictability of stock returns? *Journal of Business* 68: 309–349.
- Ferson, W.E., and A.F. Siegel. 2001. The efficient use of conditioning information in portfolios. *Journal of Finance* 56: 967–982.
- Ferson, W.E., S.R. Foerster, and D.B. Keim. 1993. General tests of latent variable models and mean variance spanning. *Journal of Finance* 48: 131–156.
- Ferson, W.E., S. Sarkissian, and T. Simin. 2003. Spurious regressions in financial economics? *Journal of Finance* 58: 1393–1414.
- Foster, D., T. Smith, and R. Whaley. 1997. Assessing goodness-of-fit of asset pricing models: the distribution of the maximal R-squared. *Journal of Finance* 52: 591–607.
- Gibbons, M.R., and W.E. Ferson. 1985. Testing asset pricing models with changing expectations and an unobservable market portfolio. *Journal of Financial Economics* 14: 217–236.
- Gordon, M. 1962. *The investment, financing and valuation of the firm*. Homewood: Irwin.
- Goyal, A., and I. Welch. 2003. Predicting the equity premium with dividend ratios. *Management Science* 49 (5).

- Hansen, L.P., and R. Hodrick. 1980. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy* 88: 829–853.
- Hansen, L.P., and R.J. Hodrick. 1983. Risk averse speculation in the forward foreign exchange market. In *An econometric analysis of linear models, exchange rates and international macroeconomics*, ed. J. Frenkel. Chicago: University of Chicago Press.
- Hansen, L.P., and S.F. Richard. 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* 55: 587–613.
- Harvey, C.R. 1989. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics* 24: 289–318.
- . 1991. The world price of covariance risk. *Journal of Finance* 46: 111–157.
- Jagannathan, R., and Z. Wang. 1996. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51: 3–54.
- Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance* 48: 65–91.
- Kim, M., C.R. Nelson, and R. Startz. 1991. Mean reversion in stock returns? a reappraisal of the statistical evidence. *Review of Economic Studies* 58: 515–528.
- Lettau, M., and S. Ludvigson. 2001a. Consumption, aggregate wealth and expected stock returns. *Journal of Finance* 56: 815–849.
- . 2001b. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109 (6): 1238–1287.
- Lo, A., and A.C. MacKinlay. 1990. Data snooping biases in tests of financial asset pricing models. *Review of Financial Studies* 3: 431–468.
- Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867–887.
- . 1980. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics* 8: 323–362.
- Paye, B., & A. Timmermann. (2003). How stable are financial prediction models? Evidence from US and International Stock Market Data. *Working Paper*, University of California at San Diego.
- Pesaran, M.H., and A. Timmermann. 1995. Predictability of stock returns: Robustness and economic significance. *Journal of Finance* 50: 1201–1228.
- Pontiff, J., and L. Schall. 1998. Book-to-market as a predictor of market returns. *Journal of Financial Economics* 49: 141–160.
- Roll, R. 1977. A critique of the asset pricing theory's tests -part 1: On past and potential testability of the theory. *Journal of Financial Economics* 4: 129–176.
- . 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance* 39: 1127–1140.
- Scholes, M., and J. Williams. 1977. Estimating beta from nonsynchronous data. *Journal of Financial Economics* 5: 309–327.
- Schwert, G.W. 2003. Anomalies and market efficiency. In *Handbook of the Economics of Finance*, ed. G.M. Constantinides, M. Harris, and R.M. Stulz. Elsevier: North Holland.
- Shanken, J. 1990. Intertemporal asset pricing: an empirical investigation. *Journal of Econometrics* 45: 99–120.
- Sharpe, W.F. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- . 1984. Factor models, CAPM's and the APT. *Journal of Portfolio Management* 11: 21–25.
- Simin, T. (2002). The (poor) predictive performance of asset pricing models. *Working Paper*, The Pennsylvania State University.
- Solnik, B. 1993. The unconditional performance of international asset allocation strategies using conditioning information. *Journal of Empirical Finance* 1: 33–55.
- Stambaugh, R.S. 1999. Predictive regressions. *Journal of Financial Economics* 54: 315–421.
- Wheatley, S. 1989. A critique of latent variable tests of asset pricing models. *Journal of Financial Economics* 23: 325–338.

Conditional Performance Evaluation

11

Wayne E. Ferson

Abstract

Measures for evaluating the performance of a mutual fund or other managed portfolio are interpreted as the difference between the average return of the fund and that of an appropriate benchmark portfolio. Traditional measures use a fixed benchmark to match the average risk of the fund. Conditional performance measures use a dynamic strategy as the benchmark, matching the fund's risk dynamics. The logic of this approach is explained, the models are described and the empirical evidence is reviewed.

Keywords

Alpha · Benchmark portfolio · Conditional alpha · Conditional beta · Investment performance · Market efficiency · Market timing · Mutual funds · Pension funds · Portfolio weights · Risk dynamics · Security selection · Stochastic discount factor

11.1 Conditional Performance Evaluation

Conditional Performance Evaluation is a collection of empirical approaches for measuring the investment performance of portfolio managers, adjusting for the risks and other characteristics of their portfolios. A central goal of performance evaluation in general, is to identify those managers who possess investment information or skills superior to that of the investing public, and who use the advantage to achieve superior portfolio returns. Just as important, we would like to identify and avoid those managers with poor performance. Since the risks and expected returns of financial assets are related, it is important to adjust for the risks taken by a portfolio manager in evaluating the returns. In order to identify superior returns, some model

W. E. Ferson (✉)

University of Southern California, Los Angeles, CA, USA

e-mail: ferson@marshall.usc.edu

of “normal” investment returns is required, i.e. an asset pricing model is needed (see ► Chaps. 9, “Asset Pricing Models”, and ► 10, “Conditional Asset Pricing”).

Classical measures of investment performance compare the average return of a managed portfolio to that of a “benchmark portfolio” with similar risk. For example, Jensen (1968) advocated “alpha” as a performance measure. This is the average return minus the expected return implied by the Capital Asset Pricing Model (Sharpe 1964). The CAPM implies that the expected return is a fund-specific combination of a safe asset and a broad market portfolio, and so this combination is the benchmark. Chen et al. (1987), Connor and Korajczyk (1986), and Lehmann and Modest (1987) extended this idea to multi-beta asset pricing models, where several returns are combined in the benchmark to adjust for the fund’s risk.

It is traditional to distinguish between investment ability for security selection and ability for market timing. Security selection refers to an ability to pick securities that are “undervalued” at current market prices, and which therefore may be expected to offer superior future returns. Market timing refers to an ability to switch the portfolio between stocks and bonds, anticipating which asset class will perform better in the near future. The classical performance measures are “unconditional,” in the sense that the expected returns in the model are unconditional means, estimated by past averages, and the risks are the fixed unconditional second moments of return. If expected returns and risks vary over time, the classical approach is likely to be unreliable. Ferson and Schadt (1996) showed that if the risk exposure of a managed portfolio varies predictably with the business cycle, but the manager has no superior investment ability, then a traditional approach will confuse common variation in the fund’s risk and the expected market returns with abnormal stock picking or market timing ability. “Conditional Performance Evaluation” (CPE) models the conditional expected returns and risk, attempting to account for their changes with the state of the economy, thus controlling for any common variation.

The problem of confounding variation in mutual fund risks and market returns has long been recognized (e.g. Jensen 1972; Grant 1977), but these studies tend to interpret such variation as reflecting superior information or market timing ability. A conditional approach to performance evaluation takes the view that a managed portfolio whose return can be replicated by mechanical trading, based on readily available public information, should not be judged to have superior performance. CPE is therefore consistent with a version of market efficiency, in the semi-strong form sense of Fama (1970).

In the CPE approach a fund’s return is compared with a benchmark strategy that attempts to match the fund’s risk dynamics. The benchmark strategy does this by mechanically trading, based on predetermined variables that measure the state of the economy. The performance measures, the “conditional alphas,” are the difference between a fund’s return and that of the benchmark dynamic strategy. This generalizes the classical performance measures, such as Jensen’s alpha, which compare a fund’s return with a fixed benchmark that carries the same average risk. Since CPE uses more information than traditional performance measures, it has the potential to provide more accuracy. In practice, the trading behavior of managers overlays

portfolio dynamics on the dynamic behavior of the underlying assets that they trade. For example, even if the risk of each security were fixed over time, the risk of a portfolio with time-varying weights, would be time varying. The desire to handle such dynamic behavior motivates a conditional approach. Investors may wish to understand how funds implement their investment policies dynamically over time. For example, how is a fund's bond–stock mix, market exposure, or investment style expected to react in a time of high-interest rates or market volatility? CPE is designed to provide a rich description of funds' portfolio dynamics in relation to the state of the economy.

A conditional approach to performance evaluation can accommodate whatever standard of superior information is held to be appropriate, by the choice of the lagged instruments, which are used to represent the public information. Incorporating a given set of lagged instruments, managers who trade mechanically in response to these variables get no credit for superior performance. To represent public information, much of the empirical literature to date has focused on a standard set of lagged variables. Examples include the levels of interest rates and interest rate spreads, dividend-to-price ratios, and dummy variables indicating calendar-related patterns of predictability. More recent studies expand the analysis to consider a wider range of indicators for public information about the state of the economy (e.g. Ferson and Qian 2004).

11.2 Examples

Implementations of Conditional Performance Evaluation have typically used either simple linear regression models or “stochastic discount factor” methods (see the Chapter “Asset Pricing”). Ferson and Schadt (1996) used simple linear regressions. To illustrate, let r_{mt+1} be the return on a market or benchmark index, measured in excess of a short-term Treasury return. For example, the benchmark index could be the Standard & Poor's 500, a “style” index such as “small cap growth,” or a vector of excess returns. The traditional regression for Jensen's alpha is:

$$r_{pt+1} = \alpha_{pJ} + \beta_p r_{mt+1} + \varepsilon_{pt+1}, \quad (1)$$

where r_{pt+1} is the return of the fund in excess of a short term “cash” instrument and α_{pJ} is Jensen's alpha. Ferson and Schadt (1996) proposed the conditional model:

$$r_{pt+1} = \alpha_p + \beta_o r_{mt+1} + \beta' [r_{mt+1} \otimes \mathbf{Z}_t] + u_{pt+1}, \quad (2)$$

where $\mathbf{Z}_t$ is the vector of lagged conditioning variables and α_p is the conditional alpha. The coefficient β_o is the average beta of the fund, and $\beta' \mathbf{Z}_t$ captures the time-varying part of the conditional beta. The interaction terms $[r_{mt+1} \otimes \mathbf{Z}_t]$ in the Ferson and Schadt regression model control for common movements in the fund's

“beta,” and the conditional expected benchmark return. The conditional alpha, α_p , is thus measured net of these effects.

To see more explicitly how Eq. 2 compares the fund’s return to a benchmark strategy with the same risk dynamics, recall that the excess returns are $r_{pt+1} = R_{pt+1} - R_{ft+1}$ and $r_{mt+1} = R_{mt+1} - R_{ft+1}$, where R_{ft+1} is the gross return of a risk-free asset. The benchmark strategy is to invest the fraction $\beta_0 + \beta'Z_t$ of the portfolio in the market index with return R_{mt+1} , and the fraction $1 - \beta_0 - \beta'Z_t$ in the risk-free investment. This benchmark strategy has a time-varying beta equal to $\beta_0 + \beta'Z_t$, the same as that ascribed to the fund. The conditional alpha is just the difference between the fund’s average return and the average return of the benchmark strategy.

Christopherson et al. (1998) propose a refinement of Eq. 2 to allow for a time-varying conditional alpha:

$$r_{p,t+1} = \alpha_{p0} + \alpha_{1p}'Z_t + \beta_0 r_{m,t+1} + \beta'[r_{m,t+1} \otimes Z_t] + u_{p,t+1}. \quad (3)$$

In this model, the term $\alpha_{p0} + \alpha_{1p}'Z_t$ captures the time-varying conditional alpha.

An alternative approach to Conditional Performance Evaluation uses “SDF” models, as developed by Chen and Knez (1996), Dahlquist and Soderlind (1999), Farnsworth et al. (2002), and Ferson et al. (2006). With this approach, abnormal performance is measured by the expected product of a fund’s returns and a SDF. (See ► Chap. 9, “Asset Pricing Models,” for a discussion of stochastic discount factors.) Specifying the stochastic discount factor corresponds to specifying an asset pricing model. For a given SDF, denoted by m_{t+1} , we can define a fund’s “conditional SDF alpha” as:

$$\alpha_{pt} \equiv E(m_{t+1}R_{pt+1}|Z_t) - 1, \quad (4)$$

where one dollar invested with the fund at time t returns R_{pt+1} dollars at time $t + 1$. In the case of an open-end, no-load mutual fund, we may think of R_{pt+1} as the net asset value return. More generally, if the fund generates a payoff V_{pt+1} for a cost $c_{pt} > 0$, the gross return is $R_{pt+1} \equiv V_{pt+1}/c_{pt}$. A SDF is said to price the vector of underlying primitive assets with returns $\mathbf{R}_{t+1}$ if their gross returns satisfy the equation $E_t\{m_{t+1}\mathbf{R}_{t+1}\} = \mathbf{1}$.

If the SDF prices the primitive assets, α_{pt} will be zero when the fund (costlessly) forms a portfolio of the primitive assets, provided the portfolio strategy uses only the public information at time t . In that case $R_{p,t+1} = \mathbf{x}(Z_t)'\mathbf{R}_{t+1}$, where $\mathbf{x}(Z_t)$ is the portfolio weight vector. Then Eq. 3 implies that $\alpha_{pt} = [E(m_{t+1}\mathbf{x}(Z_t)'\mathbf{R}_{t+1}|Z_t)] - 1 = \mathbf{x}(Z_t)'[E(m_{t+1}\mathbf{R}_{t+1}|Z_t)] - 1 = \mathbf{x}(Z_t)'\mathbf{1} - 1 = 0$.

When the SDF alpha of a fund is not zero, this is interpreted to indicate “abnormal” performance relative to the model that provides the specification of m_{t+1} . The economic intuition is simple when $m_{t+1} = \rho u'(C_{t+1})/u'(C_t)$ in the consumer choice problem: Maximize the expected utility function $E_t\{\Sigma_j \geq 0 \rho^j u(C_{t+j})\}$. Then, the condition $E_t\{m_{t+1}\mathbf{R}_{t+1}\} = \mathbf{1}$ is the necessary first-order condition of the

maximization. If the consumer–investor in this problem can invest in a fund with a given SDF alpha, the consumer–investor would wish to hold more of the fund with $\alpha_{pt} > 0$, and less of the fund with $\alpha_{pt} < 0$.

11.3 Conditional Market Timing

A classical market timing regression, when there is no conditioning information, is the quadratic regression:

$$r_{pt+1} = a_p + b_p r_{mt+1} + \gamma_{tmu} [r_{m,t+1}]^2 + v_{pt+1}. \quad (5)$$

Treynor and Mazuy (1966) argue that $\gamma_{tmu} > 0$ indicates market timing ability. The logic is that a market timing manager will generate a return that has a convex relation to the market. When the market is up, the fund will be up by a disproportionate amount. When the market is down, the fund will be down by a lesser amount. However, a convex relation may arise for a number of other reasons. Chen et al. (2005) provide an analysis of various nonlinear effects unrelated to true-timing ability. One of these is common time variation in the fund’s risk and the expected market return, due to public information on the state of the economy. In a market timing context, the goal of conditional performance evaluation is to distinguish timing ability that merely reflects publicly available information, from timing based on better information. We may call such informed timing ability “conditional market timing.”

Admati et al. (1986) describe a model in which a manager with constant absolute risk aversion in a normally distributed world, observes at time t a private signal equal to the future market return plus noise, $r_{mt+1} + \eta$. The manager’s response is to change the portfolio beta as a linear function of the signal. They show that the γ_{tmu} coefficient in regression in Eq. 5 is positive, if the manager increases market exposure when the signal about the future market return is favorable. Bhattacharya and Pfleiderer (1983), and Lee and Rahman (1990) show how to use the squared residuals of the regression to separate the manager’s risk aversion from the signal quality, measured by its correlation with the market return.

In a conditional model, the part of the correlation of fund betas with the future market return, which can be attributed to the public information, is not considered to reflect conditional market timing ability. Ferson and Schadt (1996) developed a conditional version of the Treynor–Mazuy regression:

$$r_{pt+1} = a_p + b_p r_{mt+1} + C_p' (Z_t r_{mt+1}) + \gamma_{tmc} [r_{m,t+1}]^2 + v_{pt+1}, \quad (6)$$

where the coefficient vector C_p captures the linear response of the manager’s beta to the public information, Z_t . The term $C_p' (Z_t r_{mt+1})$ controls the public information effect, which would bias the coefficients in the original Treynor–Mazuy model. The

coefficient γ_{mc} measures the sensitivity of the manager's beta to the private market timing signal, purged of the effects of the public information.

Merton and Henriksson (1981) and Henriksson (1984) described an alternative model of market timing in which the quadratic term in Eq. 5 is replaced by an option payoff, $\max(0, r_{m,t+1})$. This reflects the idea that market timers may be thought of as delivering (attractively priced) put options on the market index. Ferson and Schadt (1996) developed a conditional version of this model as well.

Becker et al. (1999) developed conditional market timing models with explicit performance benchmarks. In this case, managers maximize the utility of their portfolio returns in excess of a benchmark portfolio return. The model allows separate identification of the manager's risk aversion and skill, as measured by the signal quality. Performance benchmarks often represent an important component of managers' incentive systems, but they have been controversial, both in practice and in the academic literature. Starks (1987), Grinblatt and Titman (1989a, b), and Admati and Pfleiderer (1997) argue that benchmarks don't properly align managers' incentives with those of the investors in the fund. Carpenter et al. (2000) provide a theoretical justification of benchmarks, used in combination with investment restrictions.

11.4 Conditional Weight-Based Performance Measures

Returns-based measures of performance compare the return earned by the fund with a benchmark return over the evaluation period. The benchmark is designed to control for risk, and it may also control for style, investment constraints, and other factors. The manager who performs better than the benchmark has a positive "alpha." In some situations, information on the manager's investment positions or portfolio weights is also available. In these situations, weight-based measures of performance are attractive. With weight-based measures the manager's choices are directly analyzed for evidence of superior ability. The idea is that a manager, who increases the fund's exposure to a security or asset class before it performs well, or who avoids "losers" ahead of time, is seen to add investment value.

Cornell (1979) was among the first to propose the usage of portfolio weights to measure the performance of trading strategies. Copeland and Mayers (1982) modify Cornell's measure and use it to analyze Value Line rankings. Grinblatt and Titman (1993) proposed a weight-based measure of mutual fund performance. A number of studies have used the Grinblatt and Titman measure, including Grinblatt and Titman (1989a, b); Grinblatt et al. (1995); Zheng (1999); and Wermers (1997). These studies combine portfolio weights with unconditional moments of returns to measure performance.

Ferson and Khang (2002) consider conditioning information in weight-based measures of performance. The idea is similar to that of conditional, returns-based measures. Any predictive ability in a manager's portfolio weights that merely reflects the lagged, public information is not considered to represent superior ability. By using lagged instruments and portfolio weight data, conditional weight-based measures should provide more precision in measuring performance.

The use of portfolio weights may be especially important in a conditional setting. When expected returns are time varying and managers trade between return observation dates, returns-based approaches are likely to be biased. Even conditional returns-based methods are affected. This bias, which Ferson and Khang call the “interim trading bias,” can be avoided by using portfolio weights in a conditional setting.

The following stylized example illustrates the idea. Suppose that returns can only be measured over two periods, but a manager trades each period. The manager has neutral performance, but the portfolio weights for the second period can react to public information at the middle date. By assumption, merely reacting to public information does not require superior ability. You have to trade “smarter” than the general public to generate superior performance. If returns were independent over time there would be no interim trading bias, because there would be no information at the middle date about the second-period return.

Suppose that a terrorist event at the middle date increases market volatility in the second period, and the manager responds by shifting into cash at the middle date. If only two-period returns can be measured and evaluated, the manager’s strategy would appear to have partially anticipated the higher volatility. For example, the fund’s two-period market exposure would reflect some average of the before- and after-event positions. Measured from the beginning of the first period, the portfolio would appear to partially “time” the volatility increasing event because of the move into cash. A returns-based measure over the two periods will detect this as superior information.

In this example, since only two-period returns can be measured and evaluated, a Conditional Weight-based Measure (CWM) would examine the ability of the manager’s choices at the beginning of the first period to predict the subsequent two-period returns. To record abnormal ability under the CWM, the manager would have to anticipate the higher volatility and adjust the portfolio prior to the event. If the manager has no information beyond the public information, the CWM is zero. The ability of the manager to trade at the middle period thus creates no interim trading bias in a CWM.

The CWM is the conditional covariance between future returns and portfolio weights, summed across the asset holdings:

$$\text{CWM} = E \left\{ \sum_j w_j(Z, S) [r_j - E(r_j|Z)] Z \right\}. \quad (7)$$

The symbol $w_j(Z, S)$ denotes the portfolio weight in asset j at the beginning of the period and $r_j - E(r_j|Z)$ denotes the unexpected or abnormal excess return. The expectation is taken from the perspective of an investor, who only sees the public information Z at the beginning of the period. As viewed by an investor with this information, the sum of the conditional covariances between the weights, measured at the end of December, and the subsequent abnormal returns for the securities in the first quarter, is positive for a manager with superior information, S . If the manager has no superior information, S , then the covariance is zero.

It is important to recognize that weight-based measures do not avoid the issue of specifying a performance benchmark. For example, Eq. 7 can also be written as

$$\text{CWM} = E\{\Sigma_j w_j(Z, S) r_j - \Sigma_j E(w_j(Z, S)|Z) r_j\}. \quad (8)$$

This shows that the measure is the expected difference between the portfolio return and the return of a portfolio that uses the weights that would have been expected on the basis of the public information. The latter portfolio may be interpreted as the benchmark.

In a portfolio with abnormal performance, the covariance between the weights and subsequent abnormal returns need not be positive for every security. Consider two securities, which are highly correlated with each other. A manager with superior ability may buy one and sell the other as a result of hedging considerations. However, under certain assumptions the sum of the covariances across securities will be positive (Grinblatt and Titman 1993).

Ferson and Khang (2002) introduce an explicit “external” benchmark with weights, $w_{jb}(Z)$, which are in the public information set Z at the beginning of the period. Their empirical measure is then:

$$\text{CWM} = E\{\Sigma_j [w_j(Z, S) - w_{jb}(Z)] [r_j - E(r_j|Z)] | Z\}. \quad (9)$$

Because w_b is assumed to be known given Z , it will not affect the conditional covariance in theory. However, in practice it is desirable to measure performance relative to an external benchmark. One reason is statistical: the weights w_j may be highly persistent over time, while the deviations from benchmark are better behaved. The benchmark also helps the interpretation. Eq. 9 is the difference between the unexpected return of the fund and the unexpected return of the benchmark.

In Ferson and Khang, the benchmark at time t is formed from the actual lagged weights of the fund at $t - k$, updated using a buy-and-hold strategy. With the buy-and-hold benchmark, the measure examines the deviations between a manager’s weights and a strategy of no trading during the previous k periods. This takes the view that a manager with no information would follow a buy-and-hold strategy.

11.5 Empirical Evidence Using Conditional Performance Evaluation

There is a large body of empirical literature on the performance of mutual funds. Equity-style mutual funds have received the most attention. There are fewer studies of institutional funds such as pension funds, and a relatively small number of studies focus on fixed-income-style funds. Research on the performance of hedge funds has been accumulating rapidly over the past few years.

Traditional measures of the average abnormal performance of mutual funds, like Jensen’s alpha, are found to be negative more often than positive across many studies. For example, Jensen (1968) concluded that a typical fund has neutral

performance, only after adding back expenses. Traditional measures of market timing often find that any significant market timing ability is perversely “negative,” suggesting that investors could time the market by doing the opposite of a typical fund. Such results make little economic sense, which suggests that they may be spurious.

The first conditional performance evaluation studies, by Chen and Knez (1996), Ferson and Schadt (1996), and Ferson and Warther (1996) found that conditioning on the state of the economy is both statistically and economically significant for measuring investment performance. Ferson and Schadt (1996) find that funds’ risk exposures change in response to public information on the economy, such as the levels of interest rates and dividend yields. Using conditional models, Ferson and Schadt (1996), Kryzanowski et al. (1997), Zheng (1999), Becker et al. (1999), and Mamaysky et al. (2003) find that the distribution of mutual fund alphas shifts to the right, and is centered near zero. Farnsworth et al. (2002) use a variety of conditional SDF models to evaluate performance in a monthly sample of U.S. equity mutual funds, using a simulation approach to control for model biases. They find that the conditional performance of the average mutual fund is no worse than a hypothetical random stock-picking fund.

Ferson and Warther (1996) attribute differences between unconditional and conditional alphas to predictable flows of public money into funds. Inflows are correlated with reduced market exposure, at times when the public expects high returns, due to larger cash holdings in response to inflows at such times. In pension funds, which are not subject to high-frequency flows of public money, no overall shift in the distribution of fund alphas is found when moving to conditional models (Christopherson et al. 1998). A similar result is found for hedge funds, which often use lockup periods and notification periods to control the flows of funds (e.g. Kazemi 2003).

Henriksson (1984), Chang and Lewellen (1984), Grinblatt and Titman (1989a), Cumby and Glen (1990), Ferson and Schadt (1996), and others estimated unconditional models to assess market timing ability for equity mutual funds. They find a tendency for negative estimates of the timing coefficients. Ferson and Schadt (1996) found that this result does not occur in conditional models. Becker et al. (1999) simultaneously estimate the fund managers’ risk aversion for tracking error and the precision of the market timing signal, in a sample of more than 400 U.S. mutual funds for 1976–1994, including a subsample with explicit asset allocation objectives. The estimates suggest that U.S. equity mutual funds behave as risk averse benchmark investors, but little evidence of conditional timing ability is found. Chen (2003) finds a similar result in a sample of hedge funds, using a variety of market indexes. Jiang (2003) presents a nonparametric test of mutual fund timing ability, and again finds no evidence of ability after the effect of lagged public information variables is accounted for. Thus, controlling for public information variables, there seems to be little evidence that mutual funds have conditional timing ability for the level of the market return.

Busse (1999) asks whether fund returns contain information about market volatility. He finds evidence using daily data that funds may shift their market exposures

in response to changes in second moments. Laplante (2003) presents a model of market timing funds that accommodates timing in response to signals about both the first and second moments of return. Given the prevalence of market timing funds and the dearth of evidence that such funds can time the first moments of returns, further research on the higher moments is clearly warranted.

Ferson and Khang (2002) study the conditional weight-based approach to measuring performance. Using a sample of equity pension fund managers (1985–1994), they find that the traditional returns-based alphas of the funds are positive, consistent with previous studies of pension fund performance. However, these alphas are smaller than the potential effects of interim trading bias. The conditional weight-based measures indicate that the pension funds have neutral performance.

In summary, conditional performance measures are superior to traditional measures, both on theoretical and statistical grounds. Conditional measures eliminate the perverse, negative timing coefficients often observed with unconditional measures, and in some cases are found to deliver more precise performance measures. Overall, the empirical evidence based on conditional performance measures suggests that abnormal fund performance, after controlling for public information, is rare.

References

- Admati, A., and P. Pfleiderer. 1997. Performance benchmarks: Does it all add up? *Journal of Business* 70 (3): 323–350.
- Admati, A., S. Bhattacharya, S. Ross, and P. Pfleiderer. 1986. On timing and selectivity. *Journal of Finance* 41: 715–730.
- Becker, C., W. Ferson, D. Myers, and M. Schill. 1999. Conditional market timing with benchmark investors. *Journal of Financial Economics* 52: 119–148.
- Bhattacharya, S., and P. Pfleiderer. 1983. A note on performance evaluation. Technical Report 714, Stanford University Graduate School of Business.
- Busse, J. 1999. Volatility timing in mutual funds: Evidence from daily returns. *Review of Financial Studies* 12 (5): 1009–1041.
- Carpenter, J., Dybvig, P.H., and Farnsworth, H. 2000. *Portfolio performance and agency*. Working Paper, Washington University, St Louis.
- Chang, E.C., and W.G. Lewellen. 1984. Market timing and mutual fund investment performance. *Journal of Business* 57: 55–72.
- Chen, Y. 2003. On conditional market timing of hedge fund managers. Working Paper, Boston College.
- Chen, Z., and P.J. Knez. 1996. Portfolio performance measurement: theory and applications. *Review of Financial Studies* 9: 511–556.
- Chen, N-fu, T. Copeland, and D. Mayers. 1987. A comparison of single and multifactor performance methodologies. *Journal of Financial and Quantitative Analysis* 224: 1–17.
- Chen, Y., Ferson, F., and Peters, H. 2005. Measuring the timing ability of fixed-income mutual funds. Working Paper, Boston College.
- Christopherson, J.A., W. Ferson, and D.A. Glassman. 1998. Conditioning manager alpha on economic information: Another look at the persistence of performance. *Review of Financial Studies* 11: 111–142.

- Connor, G., and R. Korajczyk. 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* 15: 373–394.
- Copeland, T.E., and M. Mayers. 1982. The value line enigma (1965–1978): A case study of performance evaluation issues. *Journal of Financial Economics* 10: 289–321.
- Cornell, B. 1979. Asymmetric information and portfolio performance measurement. *Journal of Financial Economics* 7: 381–390.
- Cumby, R., and J. Glen. 1990. Evaluating the performance of international mutual funds. *Journal of Finance* 45: 497–521.
- Dahlquist, M., and P. Soderlind. 1999. Evaluating portfolio performance with stochastic discount factors. *Journal of Business* 72: 347–384.
- Fama, E.F. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25: 383–417.
- Farnsworth, H., W. Ferson, D. Jackson, and S. Todd. 2002. Performance evaluation with stochastic discount factors. *Journal of Business* 75: 473–504.
- Ferson, W.E., and K. Khang. 2002. Conditional performance measurement using portfolio weights: evidence for pension funds. *Journal of Financial Economics* 65: 249–282.
- Ferson, W., and Qian, M. 2004. *Conditional performance evaluation revisited*. Research Foundation Monograph. Charlottesville: CFA Institute.
- Ferson, W., and R. Schadt. 1996. Measuring fund strategy and performance in changing economic conditions. *Journal of Finance* 51: 425–462.
- Ferson, W., and V.A. Warther. 1996. Evaluating fund performance in a dynamic market. *Financial Analysts Journal* 52 (6): 20–28.
- Ferson, W., T. Henry, and D. Kisgen. 2006. Evaluating government bond fund performance with stochastic discount factors. *Review of Financial Studies*. (forthcoming).
- Grant, D. 1977. Portfolio performance and the “cost” of timing decisions. *Journal of Finance* 32: 837–846.
- Grinblatt, M., and S. Titman. 1989a. Portfolio performance evaluation: Old issues and new insights. *Review of Financial Studies* 2: 393–416.
- . 1989b. Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business* 62: 393–416.
- . 1993. Performance measurement without benchmarks: An examination of mutual fund returns. *Journal of Business* 60: 97–112.
- Grinblatt, M., S. Titman, and R. Wermers. 1995. Momentum strategies, portfolio performance and herding: a study of mutual fund behavior. *American Economic Review* 85: 1088–1105.
- Henriksson, R.D. 1984. Market timing and mutual fund performance: An empirical investigation. *Journal of Business* 57: 73–96.
- Jensen, M. 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23: 389–346.
- , ed. 1972. *Studies in the theory of capital markets*. New York: Praeger Publishers.
- Jiang, W. 2003. A nonparametric test of market timing. *Journal of Empirical Finance* 10: 399–425.
- Kazemi, H. 2003. Conditional performance of hedge funds. Working Paper, University of Massachusetts at Amherst.
- Kryzanowski, L., S. Lalancette, and M.C. To. 1997. Performance attribution using an apt with prespecified factors. *Journal of Financial and Quantitative Analysis* 32.
- Laplante, M. 2003. Mutual fund timing with information about volatility. Working Paper, University of Texas at Dallas.
- Lee, C.F., and S. Rahman. 1990. Market timing, selectivity and mutual fund performance: An empirical investigation. *Journal of Business* 63: 261–287.
- Lehmann, B., and D. Modest. 1987. Modest mutual fund performance evaluation: a comparison of benchmarks and benchmark comparisons. *Journal of Finance* 42: 233–265.
- Mamaysky, H., Spiegel, M. and Zhang, H. 2003. Estimating the dynamics of mutual fund alphas and betas. Working Paper, Yale School of Organization and Management.

- Merton, R.C., and R.D. Henriksson. 1981. On market timing and investment performance II: Statistical procedures for evaluating forecasting skills. *Journal of Business* 54: 513–534.
- Sharpe, W.F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- Starks, L. 1987. Performance incentive fees: An agency theoretic approach. *Journal of Financial and Quantitative Analysis* 22: 17–32.
- Treynor, J., and K. Mazuy. 1966. Can mutual funds outguess the market? *Harvard Business Review* 44: 131–136.
- Wermers, R. 1997. Momentum investment strategies of mutual funds, performance persistence, and survivorship bias. Working Paper, University of Colorado.
- Zheng, L. 1999. Is money smart? A study of mutual fund investors' fund selection ability. *Journal of Finance* 54: 901–933.



Working Capital and Cash Flow

12

Joseph E. Finnerty

Abstract

One of the everyday jobs of the treasurer is to manage the cash, and flow of funds through the organization. If the amount or receipt and collection activities are out of control, the entire firm may face bankruptcy. There is an old saying, “If you pay attention to the pennies, the dollars will take care of themselves.” In this spirit, this chapter looks at taking care of the daily amounts of cash flowing through the firm in a systematic fashion. The purpose is to understand the importance of the inter-relationships involved and to be able to measure the amount and speed of the cash flow. Once something can be measured, it can be managed.

Keywords

Accounts payable · Accounts receivable · Cash flow · Cash flow cycle · Cash management · Flow of funds · Inventory · Marketable securities · Matching principle · Working capital

12.1 Introduction

The management of cash flow is essential to the success of every enterprise, whether it be public or private. In fact, cash management is probably more critical to the success of an enterprise than making an individual sale or providing a service for a period of time. A business can lose a single customer or can suspend services for a short period without irreparable damage. However, let an imbalance in cash flow occur that forces a cash manager to miss a payroll, a debt payment, or a tax deadline and, quite possibly, the company is entirely out of business. This is a rather harsh penalty for one mistake or oversight on the part of the cash manager.

J. E. Finnerty (✉)
University of Illinois, Champaign, IL, USA
e-mail: finnerty@illinois.edu

During the 1960s and 1970s, when we were experiencing high rates of inflation and attendant high-interest rates, the idea of cash management became well accepted, and integrated into the financial function of the firm. This was caused by the high costs of idle cash balances. With the recession and attendant drop in inflation and lower interest rates during the 1980s, the management of cash was still important, but for different reasons. During this period even though rates were low, credit standards were tightened and cash became scarce. Again idle cash balances were undesirable. During the economic boom of the 1990s with the advent of the New Economy (Information), the changing economy caused the focus to shift from manufacturing and production to service and information. As these changes took place, the silo approach to cash management as part of the traditional treasury function shifted to a totally integrated approach focusing on creation of shareholder value. Cash management became the development and implementation of integrated financial strategies for the entire organization. During the recession (2000–2003), the economy faced low-interest rates and at the same time credit standards are tightening in the context of the new information economy, thus giving reasons for paying close attention to cash balances. In October 2004, the Federal Regulations of the U.S. Banking System for the first time were allowing all check payments to be processed electronically. This change has sped up the flow of cash through the system, and made it more important for treasurers to keep track of cash balances.

The purpose of cash flow and working capital management has become an indispensable part of the entire organization. The objective is no longer to maximize cash flow or minimize idle cash but rather to ensure the availability of cash in the form and amount required for the continuing operation of the enterprise and to ensure an addition to shareholder value.

Standard texts on the subject include: Gallinger and Healey (1991) and Maness and Zietlow (1998).

12.2 Definitions

12.2.1 Working Capital

The following terms are those more commonly used in connection with working capital.

Working capital is the dollar amount or the total of a firm's current assets. Current assets include cash, marketable securities, investments, accounts receivable, and inventories. These assets are considered liquid because they can be converted into cash within a year. The dollar amount of these assets varies from time to time because of seasonal variations in sales and cyclical variations in general business conditions. Hence, the level of working capital held by a company is not constant.

Working capital can be thought of consisting two parts – permanent and temporary. Permanent working capital is the dollar amount of working capital that remains fairly constant over time, regardless of fluctuations in sales. Temporary working

capital is the additional assets required to meet the seasonal or cyclical variations in sales above the permanent level.

12.2.2 Working Capital Management

Working capital management is a much broader concept than working capital because it involves the management of current assets, current liabilities, and the interrelationship between them. In practice, we tend to make no distinction between the investment decisions regarding current assets and the financing decisions regarding current liabilities. In fact, quite often these two are so closely related that we talk about spontaneous financing of assets – for example, a firm buying some inventory on credit. In such a situation, both assets and liabilities are increased simultaneously thereby providing, at least in the short run, the financing for the investment.

12.2.3 Net Working Capital

Net working capital is the difference between current assets and current liabilities. It is a financial indicator that can be used in conjunction with ratios to gauge the liquidity of a company. In general, an abundance of net working capital is considered desirable because it suggests that the firm has ample liquidity to meet its short-term obligations. As we shall see, this may not always be the case. In fact, one of the objectives of cash management is to reduce excess or redundant net working capital to a minimum, and thereby reduce the cost of holding idle assets.

12.3 An Overview of Corporate Working Capital

12.3.1 Money

The subject of this chapter is cash flow, or in other words, how money moves through a business enterprise. Everyone has a general understanding of what money is and how it can be used. A simple definition of money, one used by the Federal Reserve, is: Money is made up of the currency in circulation and checking account balances. The characteristics that must be present for something to serve as money are, first, a store of value; second, ready acceptance; and third, easy transferability.

Throughout history, we have seen various things serve as money – for example, the giant stone of the Yap Islanders, the tobacco currency of early American colonists, gold, silver, shells, and even paper. The key feature that these diverse things have in common is that the participants in the economy were willing to use them for transaction purposes, or to represent the accumulation of wealth. In the new age of the internet, money has taken the form of information. There is no physical representation of value, i.e. dollar bills, credit cards, etc., but rather information with

respect to account numbers and the ability to transfer value from one account to another. Such things as digital cash, digital wallets, and virtual credit cards are being used as e-money. This new approach will have an impact on working capital with respect to time and costs. From this understanding of the function that money serves, we can move to a much more sophisticated concept – that of the flow of funds.

When financial managers talk of the flow of funds or working capital, they are referring to the fact that money as we know it – corporate cash checking account and e-money – is actually increasing or decreasing as a result of management actions or decisions. However, they are also referring to factors or accounts that are not really money, but which serve as close substitutes. Such things as inventory, accounts receivable, financial instruments, and other types of marketable securities are all affected by economic or corporate activity. As these accounts change, the ultimate effect is a change in the level of corporate cash. But before these so-called near monies are actually turned into money, we can keep track of them by observing the corporate flow of funds.

12.3.2 Cash Management

Maximum cash generation is usually the primary objective of the financial manager. This objective is based on the assumption that any business is only as sound as the management of its cash flow. However, cash flow management is not an isolated task in the normal operation of a business. Instead, managing cash flow means being deeply involved with every aspect of business operations. Consequently, any and all management effort must be directed to at least satisfying cash flow requirements while managers try to achieve the other objectives of the company. To be more specific, cash flow must be considered to achieve survival, profitability, growth, creation of shareholder value, and finally, the efficient use of corporate resources.

No one objective or goal predominates at all times. The goals are interrelated to such a degree that it is in the best interest of management to work toward attaining all the goals simultaneously. At any given time, priorities may vary as to which objective is most crucial, but all of them must ultimately be achieved to run a successful enterprise.

Keynes' famous statement "In the long run, we're all dead" does not necessarily apply to the corporate form of business. Survival becomes one of the primary objectives for any business. Temporary illiquidity, or lack of money, or financial resources may lead to suspending payments of corporate obligations. As long as creditors accept deferred or rescheduled payments, the short-run problems may be worked out and the business may survive. The ultimate threat of creditors is to drive a business into bankruptcy, which is in effect the admission by management that the cash values from dissolving the business is worth more than trying to keep the business going.

From the cash flow manager's perspective, the desire for survival demands that the firm be managed in such a way as to guarantee the maximum cash flow possible. Thus, management seeks to convert the company's investment in inventories and

receivables into cash as quickly as possible. Remember that this desire to speed up cash inflow must be balanced against growing revenues, increasing profits, and the creation of shareholder wealth. In the extreme case, a company could make every product on an order basis and demand cash payment. This would eliminate inventories and receivables. No doubt, the competitive structure of any industry would reduce this strategy to a very unprofitable one in short order.

Other things being equal, the higher the profits a company generates, the more successful it is in achieving its other goals. However, as a business seeks to maximize profits, it must take greater risks. As the risk increases, the need for careful cash management becomes much more important. As a business strives to become more profitable by becoming more competitive, there is a cost in terms of higher inventories, more efficient production equipment, and more liberal credit policies to encourage sales. Competitive strategies increase the firm's need for cash flow by slowing down the rate at which working capital is converted into cash and by increasing the amount of resources tied up in each of the working capital components. Indeed, the cash manager will constantly be forced to balance profitability, growth, and survival as the manager tries to ensure that the company not only has sufficient funds but also uses those funds in an efficient fashion.

Rapid expansion in revenue and increase in market share make marketing management an exciting profession. Marketing-oriented individuals measure their success not by increased profits, but by the increase in the year's market share or by the percentage of market share a given product line has achieved. In striving for these objectives, quite often the risks of rapid growth are overlooked. The major problems begin to surface when the cash management system has not kept up with the rapid growth and its attendant increase in risk.

The efficient use of leverage is of primary importance to sustaining rapid growth. The owners of a company do not have the liquid resources to provide all of the cash necessary to finance the growth, and so external funding must be sought. Usually this external funding is in the form of debt, which increases the overall risk for the company. In and of itself, leverage is neither good nor bad. However, the misuse of leverage can place severe drains on the cash flow of a firm at a time when the company can least afford these drains.

An effective financial manager must balance the multiple objectives of the firm, and keep in mind that there are many ways to achieve these objectives, and use the firm's resources efficiently. Too much emphasis on any one of these goals can lead to very severe cash flow problems. The effective management of cash flow is necessary to achieve the multiple objectives simultaneously.

12.3.3 The Components of Working Capital

The components of working capital are the current assets listed on a firm's balance sheet – cash, marketable securities, accounts receivable, and inventory. We can envision the flow of funds through a company as the process of continuously converting one asset into another. Cash is used to buy the necessary raw material

that will be used in the production of goods and services. These goods are sold to customers. This increases accounts receivable. As customers pay their bills, accounts receivable are once again turned into cash. If there is a temporary surplus of cash, it may be used to purchase marketable securities. By holding marketable securities, a firm can earn interest on surplus funds, but can quickly convert these funds back into cash when needed. The company then repeats the cycle. The amount of funds and the speed at which the funds move from one account to another are the essential elements of cash flow management.

12.4 Flow of Funds

The flow of funds through an organization encompasses all segments of the corporation and is related to all decisions within the firm. This flow of cash, or flow of funds, is one of the main concerns of the cash manager. The flow of funds diagram below illustrates how funds flow through a company. Because the flow is circular and continuous, it is possible to start anywhere in the diagram.

12.4.1 Cash

Cash is listed first on the balance sheet because it is the most liquid shortest term asset. In the flow of funds diagram, it is at the heart of the process. A company may keep a small amount of actual currency on hand as petty cash, but this quantity of cash is usually very small compared to the company's demand deposits, checking account balances, or lines of credit. Demand deposits are the principal way in which a corporation pays its bills, both by issuing a check or electronic funds transfer.

The main problem that financial managers face is maintaining the cash account at an appropriate level. If they hold too little cash, they run the risk of not being able to pay the bills, or take advantage of opportunities that arise. On the other hand, holding too much cash is not good, because the interest that would have been earned if the funds had been properly invested is lost. The process of balancing too little versus too much cash demands most of a cash manager's attention.

12.4.2 Marketable Securities

Marketable securities are closely related to cash. In fact, they are often called cash equivalents and may be combined with cash on the company's balance sheet. Investing in marketable securities involves purchasing money market instruments. These include treasury bills, commercial paper, certificates of deposit, and other short-term investments. A ready secondary market exists for such securities because most companies regularly buy and sell them before they mature. Because such a large market exists, any company can easily sell these instruments at a price close to their true value. This is why they are called marketable securities (Fig. 1).

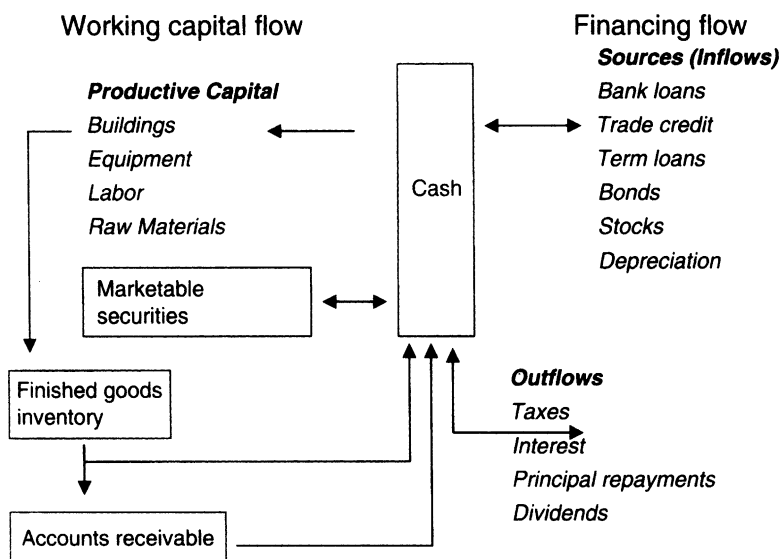


Fig. 1 Flow of funds

Most firms invest excess cash balances in marketable securities because they earn interest.

The financial manager is faced with two problems when managing the marketable securities account. First, how much money should he invest? And second, what is the appropriate maturity? When making this decision, other things being equal, the longer the maturity, the higher the yield on the investment.

12.4.3 Accounts Receivable

Accounts receivable consist of the money owed to the company by customers. Accounts receivable exist because most firms sell on credit. Customers buy now and pay later.

Accounts receivable usually constitute a very large component of a company's working capital. Thousands, even millions, of dollars can be tied up in a firm's accounts receivable. Why do companies make such a large commitment? The answer is, of course, that most companies extend credit to customers. This is primarily for marketing reasons. Customers are more willing to buy on credit, and very often competitors are willing to offer credit. In most businesses, credit terms are determined by traditional industry practice and competitive conditions. The automobile industry is a great example of using credit terms to sell cars.

Many financial managers work very closely with the marketing department to determine credit terms. This is because financial managers are responsible for obtaining the funds needed to finance accounts receivable. However, financial

managers do have some control over the level of accounts receivable by ensuring prompt billing and collection.

12.4.4 Inventories

Inventories are the physical materials that company uses to make its products, or to sell directly to its customers. Companies maintain inventories for two main reasons: first, it is more efficient and less expensive to buy from suppliers in large quantities; and second, many customers demand a wide selection of products and fast delivery. If a company is not able to offer its customers wide choice and fast delivery, it will lose sales to competitors.

We have discussed the four components of working capital – cash, marketable securities, accounts receivable, and inventory. Associated with each of these components is too small or too large.

Too much cash has an opportunity cost of forgone interest, which will hurt profitability. Too little cash may lead to a situation in which the company is unable to meet its commitments and is forced into bankruptcy despite profitable operations. An excess of funds tied up in marketable securities can lead to slower company growth because the funds have not been efficiently used for expansion. Insufficient funds in marketable securities may lead to an inadequate safety margin when cash is needed in an emergency.

Too little cash tied up in accounts receivable may indicate a noncompetitive credit policy, a business downturn, or a dwindling market for the company's products. Too large an amount in accounts receivable may indicate an overgenerous credit policy, which in turn, could lead to collection and bad-debt problems, and inefficient use of the firm's resources.

Finally, too much inventory incurs the risk of obsolescence as well as additional costs of storage, insurance, and handling. On the other hand, too little inventory may place the firm in a noncompetitive position for failing to have the products to sell when the consumer wants them.

12.4.5 The Accounting Perspective Versus the Financial Perspective

A manufacturer or wholesaler seldom generates a sale directly in exchange for cash. Instead, the firm exchanges a product for the IOU of the customer according to predetermined selling terms. When a company purchases inventory, the cash payment typically follows the actual receipt of the inventory by 30 days.

From the accounting perspective, no distinction is made between an actual transaction and a cash transaction. Thus, on the seller's side, a transaction requires a record of the sale on the day it occurs, even though no cash actually changes hands. The buyer's side also records the purchase, and at the same time, records an increase in inventory and accounts payable. But the actual transaction has no immediate effect on the cash account of either company. This is known as accrual accounting.

We can define accrual basis accounting as the recognition of revenue when it is earned and the recognition of expense in the period in which it is incurred, without regard to the time of receipt or payment of cash.

Financial accounting enables a manager to measure the financial performance of a firm by properly matching the firm's revenues and expenses as they occur. At the same time, however, accrual accounting does not provide the proper picture of the cash flow through the company. Although it is well known that corporate managers seek to maximize profit and maintain corporate liquidity, accounting theory focuses almost exclusively on measuring and reporting profitability. Any use of an accrual accounting system to measure cash flow is just as foolhardy as the use of a cash budget to measure profitability. These diverse systems were designed to measure different types of activity.

One very important financial statement presented by firms is the cash flow statement. Basically it provides information about cash flows from three areas of firm activity: (1) cash flow from operating activity, (2) cash flow from investing activity, and (3) cash flow from financing activity. The cash flow from operations is merely the reported net income plus a minus the change in net working capital plus depreciation. The cash flow from investing activity is any purchase or sale of fixed assets needs a plant, equipment or land, and the cash flow from financing activities considers the issuance of equity or debt as well as the repayment of debt, the repurchase of equity, and the payment of preferred and common dividends. Using these classifications of cash flow, we can identify "free cash flow." The firm's free cash flow is defined as cash provided by operating activities minus capital expenditures net of depositions minus preferred dividends. The amount of free cash flow available to management will allow for flexibility in making decisions about the firm's future.

It must be noted that the accounting perspective (net income) and the financial perspective (free cash flow) are very different. Each serves a different function and uses different forms of analysis to give different perspectives on a given firm's performance.

12.4.6 The Reasons for Holding Cash

As we have already said, cash is listed first on a company's balance sheet and is considered a component of working capital. Cash is made up of demand deposits and currency. Now, let's examine the reasons for holding cash.

There are three principal reasons for holding cash. First, a company needs cash for transactions. This cash is used to pay bills, wages, taxes, and meet other company obligations. We have already seen that having a positive net income does not guarantee that a company has enough cash on hand to meet all of its obligations.

The second reason for keeping a supply of cash is to have it available as a reserve. The old rule of saving for a rainy day is just as applicable for corporations as it is for individuals. Financial managers cannot predict exactly what future cash needs will be. Therefore, managers must have some cash in serve to meet unexpected needs.

The exact amount of cash held in reserve depends on the degree of uncertainty about these needs. If there is a great deal of uncertainty about day-to-day cash needs, the company will have to maintain a large cash reserve. The necessity for maintaining a large cash reserve is lessened if the company has fast, dependable, and easy access to short-term credit. For example, if a bank extends a line of credit that can be used during times of cash shortages, lower cash reserves can be maintained.

Finally, holding cash is an essential part of many lending indentures. When a firm borrows money, the lender requires certain conditions (covenants) that must be adhered to, for example, a certain level of cash must be maintained in a bank account. In order to be in legal compliance with the lending agreement, the firm must maintain a minimum level of cash or working capital

12.4.7 Investing in Marketable Securities

Most cash held in demand deposits earns no interest. Therefore, once the basic corporate needs for cash are satisfied, the financial manager should invest extra cash in the most productive manner possible. Many cash managers invest at least a part of this surplus money in marketable securities. As mentioned earlier, marketable securities earn a reasonable rate of return and offer the advantage of being quickly convertible into cash.

There are four criteria that should be considered when evaluating marketable securities: safety, marketability or liquidity, yield, and taxability. Safety refers to the probability that the full principal will be returned without any loss. Financial managers require a very high degree of safety in marketable securities. Marketability refers to how quickly and easily a security can be converted into cash. This factor is especially important if the security is being held as a reserve for the cash account, because it may have to be sold on very short notice. Yield is the interest or the price appreciation received from holding the security. Some securities pay interest; some may have tax-free interest; other securities sell at a discount and pay full face value at maturity. The effect is the same as paying interest. Some securities may have tax-free interest; discount securities may be taxed at a different rate than interest paying securities; and so on. Therefore, a cash manager must be aware of the corporation's tax situation in order to select the best type of marketable security.

There are three main reasons for investing in marketable securities. One is that they act as a reserve for the cash account. In other words, marketable securities are held to meet unexpected cash needs. Therefore, as noted earlier, their marketability is very important because they may have to be sold quickly.

Securities can also be used to meet known cash outflows. Frequently the need for certain cash outflows can be predicted. One example is taxes. Every company regularly withholds taxes from employees' paychecks. This money is paid to the government on a monthly or quarterly basis. The cash manager knows the exact amount and the due dates of these payments in advance and can purchase securities that mature at the correct time.

A third reason for investing in marketable securities is that company profits benefit. Cash managers describe funds not needed for cash reserves or taxes as “free” because such funds are not constrained by specific liquidity requirements. Therefore, cash managers can invest free cash for a higher yield after considering taxes, even though such investments are less liquid and may be a bit more risky.

12.4.8 Creating an Integrated Cash Management System

There are two main benefits to be derived from a cash management program – first, incremental profits that will augment net income, and second, freed-up resources (namely cash) that can be used for other corporate purposes. Both of these benefits are worthwhile, but the most important benefit is probably an effective cash management system. Such a system will not only pay for itself but should also have a positive effect on net income.

Reviewing the cash management cycle from beginning to end is the best approach to integrating cash management with overall company planning. The objective of the review is to find all the ways (that are consistent with the firm’s other objectives) to speed up inflows and slow down outflows. The emphasis should be on evaluating all corporate functions that, from a financial executive’s standpoint, can potentially affect cash flow.

An integrated cash management analysis involves reviewing the firm’s billing and collection procedures in light of industry practice and competition. The purpose of the review is to shorten the time it takes for payments to be put to some useful purpose.

Many banks offer cash management services to both corporate and individual clients. The relationship the company has with its bank and the form and amount of bank compensation must be reviewed carefully. In addition, an in-depth review of forecasting and planning procedures must be done to ensure that management has a good understanding of the company’s cash flow cycle. Both the timing and the amounts of flows must be taken into consideration so that the firm’s short-term investment performance will produce an acceptable rate of return.

A total review of a cash management system should also look beyond cash mobilization to information mobilization. This ensures that the decision maker receives information quickly so excess funds can be invested or short-term liabilities reduced. Clearly, it makes no sense to mobilize a company’s cash if productive uses for the additional funds are not exploited.

The next step in a cash management system is to integrate it with the financial management information system. This means setting up a planning and budgeting system that identifies projected financial needs, forecasts surpluses or deficits of funds, and then makes coordinated decisions to use those funds most effectively. Systematically coordinating short- and long-term activity allows a financial executive to know at all times what is happening at the bank, in the firm’s marketable security portfolio, to the firm’s capital budget plans, and to overall corporate liquidity needs.

The benefits of using cash more efficiently are readily apparent. But if all of the financial functions are combined, the overall cost of managing such a system is reduced. Thus, from both income-generating and cost-reduction perspectives, a cash management system can be self-sustaining.

To develop and implement broader integrated systems, cash managers must take more responsibility for coordinating and working with executives in other functional areas of the firm. It is also important to review corporate policies and procedures, to determine whether there is full interaction in the cash management function.

As a result of the wide-ranging impact of cash management on the entire firm, financial executives have a more complex job than ever before. Such executives must broaden their interests and interactions while at the same time performing the traditional financial functions. For example, they have to interact with the purchasing department and with the materials management staff. They also must play a much larger role in contract negotiations. Too often, contracts are left to the legal department, and some important financial considerations may be overlooked, especially as these considerations relate to the firm's cash management policies. Costs, payments, disbursement schedules, progress payments, and other financial considerations are of concern to financial executives. They should be involved in negotiations before contracts are finalized.

The idea of taking a company-wide view rather than looking specifically at individual operations makes eminent sense. Cash managers should broaden their perspective and think of cash management as an activity that is affected by all components of the company's operations. All decisions that are made and actions taken through the company affect cash flow, and hence cash management. This includes everything from production scheduling and inventory control to marketing and credit policy – from tax policy, negotiation, accounting, and control to personnel and payroll. The effects of cash management have an impact on all areas of a company.

The simple statement that cash management speeds up inflows and slows down outflows must be put into perspective for it to be effective and useful. Emphasis should be placed on evaluating how well the organization performs in all cash management areas, and how effectively the concepts as a whole.

12.4.9 Cash Flow Cycle

The flow of funds or cash flow refers to the movement of money through the business. The time it takes for these funds to complete a full-cycle reflects the average duration that a firm's cash is invested in inventory and accounts receivable, both of which are non-earning assets. Therefore, it is in a company's best interest to keep the cash cycle as short as possible.

To see the relationship between the various accounts and the cash cycle refer to the cash cycle chart below.

The table below shows how cash, accounts receivable, inventory, and accounts payable will be affected by each of the four steps in the cash cycle (Fig. 2).

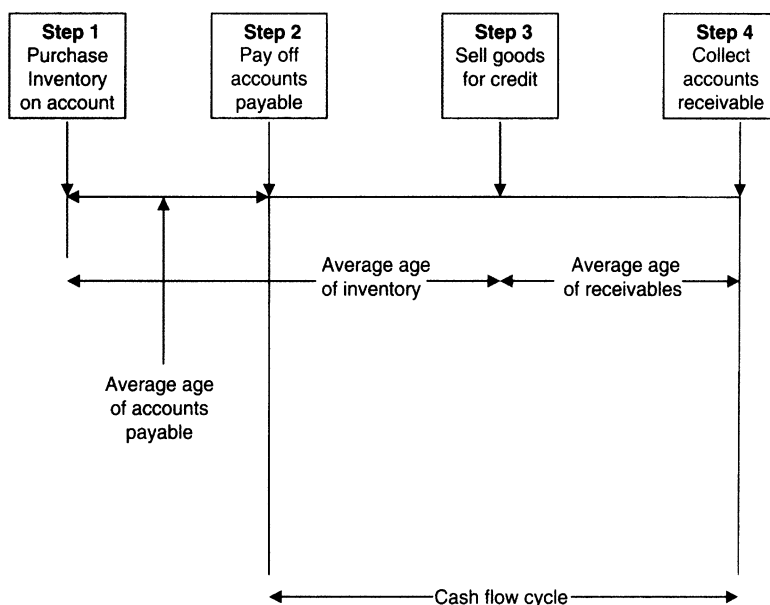


Fig. 2 Cash cycle

12.5 Calculating the Cash Flow Cycle

The cash flow cycle is defined as the average age of inventory plus the average accounts receivable less the average age of accounts payable. Or as the cash cycle chart indicates, the cash flow cycle is the average time it takes for a company to pay out cash, and receive a cash inflow that completes the transaction.

The cash flow cycle can be calculated using the following equation:

$$\begin{aligned}
 \text{Cash flow Cycle} &= \text{Average age of inventory} \\
 &\quad + \text{Average age of accounts receivable} \\
 &\quad - \text{Average age of accounts payable} \\
 &= \frac{360 \text{ Days}}{\text{Cost of goods sold/Inventory}} \\
 &\quad + \frac{\text{Accounts receivable}}{\text{Net credit sales}} \times 360 \text{ Days} \\
 &\quad - \frac{\text{Accounts payable}}{\text{Credit purchases}} \times 360 \text{ Days}
 \end{aligned}$$

To illustrate how this equation can be put to use, let us look at a Company ABC (Table 1), where:

Table 1 Changes in accounts

Account	Step 1	Step 2	Step 3	Step 4
Cash	0	—	0	+
Accounts receivable	0	0	+	—
Inventory	+	0	—	0
Accounts payable	+	—	0	0

Cost of goods sold = \$470,570

Inventory = \$345,420

Accounts receivable = \$70,820

Net credit sales = \$575,460

Accounts payable = \$26,890

Credit purchases = \$352,927

Substituting these figures in our equation, we calculate the cash flow cycle of ABC Company as:

$$\begin{aligned}
 \text{Cash flow cycle} &= \frac{360}{470,475/345,420} + \frac{70,820}{575,460} \times 360 - \frac{26,890}{352,927} \times 360 \\
 &= 265 + 44 - 27 \\
 &= 282 \text{ days}
 \end{aligned}$$

In the above example, ABC's cash cycle is 282 days. It is not possible to decide at this point if this amount of time is too long or too short because it represents a lot of factors that need to be considered. Calculating your company's cash flow cycle is the starting point of any analysis that you will have to perform to answer the question: Is cash being effectively managed in my company?

12.6 The Matching Principle

One of the fundamental principles of finance is matching the cash inflows from assets with the cash outflows from their respective sources of financing. The technique of hedging can be used when trying to accomplish this objective.

Financing temporary current assets with short-term sources of funds and financing fixed assets and permanent current assets with long-term sources of funds illustrates how the matching principle is put to work. The basic strategy of the perfect hedge is to match the expected inflows and outflows of funds. This is fundamentally sound financing because the inflows of funds from the sale of assets are being used to repay the loans that financed these acquisitions. When the cash inflow is in excess of the required cash outflow, the situation is considered to be more

conservative than the opposite case in which the cash outflow is greater than the cash inflow. This imbalance must be met by rolling over short-term financing or seeking other sources of funds. This is considered to be an aggressive approach.

The company, DEF, demonstrates the trade-off that exists between risk and return when using different approaches to the matching principle. The trade-off should be kept in mind when a company is considering a change in its sources of funds in response to changing conditions. The aggressive approach should be used when firms are expanding their working capital during the recovery and prosperity phases of the business cycle. Alternatively, during the recessionary phase, a more conservative approach may be more appropriate.

12.7 A Conservative Versus an Aggressive Approach to the Matching Concept

The figures below illustrate the results DEF Company would achieve by employing an aggressive approach or a conservative approach to matching its cash inflows with its cash outflows (Table 2).

Table 2 DEF Company (dollars in millions)

Assets			
Current assets	\$100		
Fixed assets	100		
Total assets	\$200		
Liabilities			
		Conservative	Aggressive
Short-term liabilities (at 7%)		\$25	\$100
Long-term debt (at 12%)		125	50
Equity		50	50
Total liabilities plus equity		\$200	\$200
Income statement			
Earnings before interest and taxes		\$50.00	\$50.00
Less			
Interest		(16.75)	(13.00)
Taxes (at 40%)		(13.30)	(14.80)
Net income		\$19.95	\$22.20
Current ratio	$\left(\frac{\text{Current assets}}{\text{Current liabilities}} \right)$	4.0	1.0
Net working capital	$\left(\text{Current assets} - \right. \\ \left. \text{Current liabilities} \right)$	\$75.00	\$0.00
Rate of return on equity	$\left(\frac{\text{Net income}}{\text{Equity}} \right)$	39.9%	44.4%

12.8 Summary

The purpose of this chapter is to put working capital into the proper perspective for managers concerned with managing the cash flow of their firms. They must realize that planning and managing cash flow are more than just managing the cash account. Although the cash account is one of the major assets that affects cash flow, other current assets and current liabilities, and quite often long-term assets and financing, it also has an impact on the cash cycle of the firm.

The concept of working capital and management's philosophy of trying to maintain it at a particular level are also important. If management is aggressive, it may take one approach to the matching principle that will have direct impact on the cash flow planning process. If management tends to be more conservative, other options may be available. Above all, when we are dealing with the cash flow planning process, it must be remembered that we are involved with a very dynamic situation that is closely related to the character of the decision maker. Therefore, given the exact same situation, two different managers can reach satisfactory solutions that may be entirely different from one another.

References

- Gallinger, G.W., and P.B. Healey. 1991. *Liquidity analysis and management*. 2nd ed. Reading: Addison-Wesley.
- Maness, T.S., and J.T. Zietlow. 1998. *Short term financial management*. Fort Worth: The Dryden Press.



Evaluating Fund Performance Within the Stochastic Discount Factor Framework

13

Jonathan Fletcher

Abstract

Glosten and Jagannathan (J Empir Financ 1:133–160, 1994) and Chen and Knez (Rev Financ Stud 9:511–555, 1996) pioneered the stochastic discount factor approach to fund performance. A major attraction of this approach is that it has been widely used to evaluate the performance of a wide range of managed funds such as mutual funds, hedge funds, and venture capital funds among others. In this chapter, I present an overview of the use of the stochastic discount approach to evaluate the performance of managed funds, and the estimation issues involved in this approach.

Keywords

Stochastic discount factor · Managed funds · GMM

JEL Classification

G11 · G12

J. Fletcher (✉)

Department of Accounting and Finance, University of Strathclyde, Glasgow, UK

e-mail: j.fletcher@strath.ac.uk

13.1 Introduction

Glosten and Jagannathan (1994) and Chen and Knez (1996) develop the stochastic discount factor approach to fund performance.¹ Major attractions of the stochastic discount factor approach are that most asset pricing models can be written as a candidate model of the stochastic discount factor, and the same performance measure can be used for any type of managed funds. This is of particular benefit when evaluating the performance of funds that engage in highly nonlinear and non-normal trading strategies such as hedge funds (Almeida et al. (2020)). The stochastic discount factor approach has been used in a large number of studies. A partial list includes Dahlquist and Soderlind (1999), Farnsworth et al. (2002), Ferson et al. (2006), Ahn et al. (2009a), Korteweg and Nagel (2016), Li et al. (2016), Ferson and Mo (2016), Chretien and Kammoun (2017, 2018, 2019), Almeida et al. (2020), and Gredil et al. (2020) among others.

In this chapter, I present an overview of the stochastic discount factor approach to evaluate fund performance and discuss a number of estimation issues in using the stochastic discount factor approach to evaluate fund performance. The chapter is organized as follows. Section 13.2 provides an overview of the stochastic discount factor approach to fund performance. Section 13.3 discusses estimation issues in evaluating fund performance using the stochastic discount factor approach, where linear factor models are used as they are the most commonly used models in evaluating fund performance. The final section concludes.

13.2 Evaluating Performance

Ross (1978), Harrison and Kreps (1979), and Hansen and Richard (1987), among others, show that if the law of one price (LOP)² holds in financial markets, then there exists a stochastic discount factor m_{t+1} such that:

$$p_{it} = E_t(m_{t+1}x_{it+1}), \text{ for } i = 1, \dots, N \quad (1)$$

where x_{it+1} is the payoff of asset i at time $t+1$, p_{it} is the cost of asset i at time t , N is the number of test assets, and E_t is the expected value conditional on the information available to investors at time t . Where the payoff of asset i is equal to the gross return of asset i ($1 + R_{it+1}$), the cost equals one. Where the payoff of asset i is an excess return (r_{it+1}), the cost equals zero. Equation (1) states that the expected value of the risk-adjusted payoff of asset i at time $t+1$ conditional on information available at time t has a cost equal to p_{it} . Under the assumption of no arbitrage (NA), m_{t+1} will

¹Cochrane (2005) and Ferson (2019) provide a review of the stochastic discount factor approach to asset pricing. Ferson (2010, 2013) provides excellent reviews of the stochastic discount factor approach to fund performance.

²The LOP states that two assets with identical payoffs will have the same price.

be positive in every state of nature (Cochrane (2005)). In complete markets, m_{t+1} will be unique.

Equation (1) can be written in terms of excess returns as:

$$E_t(m_{t+1}r_{it+1}) = 0 \quad (2)$$

Where there is no conditioning information, Eq. (2) implies that:

$$E(m_{t+1}r_{it+1}) = 0 \quad (3)$$

Equation (3) can be rearranged using the definition of covariance as:

$$E(r_{it+1}) = -\text{cov}(m_{t+1}r_{it+1})/E(m_{t+1}) \quad (4)$$

Equation (4) states that the expected excess return of asset i depends upon the covariance between the stochastic discount factor and excess returns of asset i ($\text{cov}(m_{t+1}r_{it+1})$). The $\text{cov}(m_{t+1}r_{it+1})$ captures the risk adjustment implied by the stochastic discount factor model. Assets with a negative covariance with the stochastic discount factor have higher expected excess returns.

Chen and Knez (1996) present a general framework to evaluate fund performance using a candidate model of the stochastic discount factor. Chen and Knez (1996) show that the unconditional performance of the fund can be written as:

$$\alpha_p = E(m_{t+1}r_{pt+1}) \quad (5)$$

where α_p is the performance of the fund, and r_{pt+1} is the excess return of the fund in period $t+1$. The performance of the fund (α_p) measures the difference between the expected risk-adjusted excess return of the fund and its cost (0). If performance is positive (negative), the fund offers a higher (lower) risk-adjusted excess returns than expected which signifies superior (inferior) performance.

The conditional performance of the fund is given by:

$$\alpha_{pt} = E_t(m_{t+1}r_{pt+1}) \quad (6)$$

The conditional performance of the fund α_{pt} measures the difference between the expected risk-adjusted excess return of the fund at time $t+1$ conditional on information available at time t minus its cost. The conditional performance of the fund varies over time as a function of conditioning information. If we are evaluating fund performance using the gross returns of the fund, then we would subtract 1 from Eqs. (5) and (6).

What is the goal of a performance measure? Chen and Knez (1996) point out that a performance measure seeks to measure the value added by a professional portfolio manager. Does the portfolio manager enlarge the investment opportunity set of investors? Aragon and Ferson (2008) point out that the issue of what the performance measure captures depends upon how costs are treated. Aragon and Ferson define performance ability when fund performance is evaluated gross of all costs and

expenses by the fund. Value added is defined when fund performance is evaluated net of all costs and expenses by the fund.

The question of performance ability or value added can be addressed by using unconditional performance measures or conditional performance measures. An unconditional framework assumes that the investment opportunity set of uninformed investors is spanned by passive trading strategies in the N test assets. A conditional framework allows for uninformed investors to follow dynamic trading strategies in the N test assets based on publicly available information. The use of the conditional performance framework provides a greater challenge to the portfolio manager. Portfolio managers who trade on the basis of public information will not be rewarded for superior performance within the conditional framework. In contrast, the portfolio manager can be rewarded superior performance by trading on public information in an unconditional framework (see Ferson and Schadt (1996), Ferson and Khang (2002) for further discussion).

Chen and Knez (1996) define four conditions that are required for admissible performance measures to exist. The most important condition is that any trading strategy that can be achieved by an uninformed investor should be given zero performance. Chen and Knez show that admissible performance measures exist when the LOP holds in financial markets. Admissible performance measures are equivalent to the existence of admissible stochastic discount factors that can price the set of assets that the uninformed investor can trade in. When financial markets satisfy the NA restriction and the fund is evaluated by a NA admissible stochastic discount factor, such performance measures are defined as NA admissible performance measures.

Chen and Knez (1996) and Farnsworth et al. (2002) show that portfolio managers who trade in the N test assets (without superior information and trading costs) will be assigned zero performance by all admissible performance measures. For admissible unconditional measures, the performance of the fund will be zero when the manager does not trade on any information and for admissible conditional measures, the performance of the fund will be zero when the manager only trades on public information.

The ambiguity in fund performance for admissible performance measures is when the portfolio manager's return cannot be perfectly replicated by the test assets. In this situation, the fund performance will be sensitive to the admissible stochastic discount factor model used. Chen and Knez (1996) show that there are different admissible stochastic discount factor models that price the test assets correctly but can give the same fund positive and negative performance.³

If we add in the NA restriction, Chen and Knez (1996) show that fund performance is still sensitive to the NA admissible stochastic discount factor model used. This analysis is extended further by Ahn et al. (2009a) who derive the NA performance bounds of a managed fund given by all NA admissible stochastic discount factors. The empirical analysis in Ahn et al. show that these NA bounds are wide for

³This issue is related to the benchmark problem in fund performance using traditional performance measures such as Roll (1978), Lehmann and Modest (1987), and Chan et al. (2009) among others. However, these studies are including benchmark models that result in inadmissible stochastic discount factors.

US mutual funds. The one positive result from Chen and Knez is that if a NA admissible stochastic discount factor assigns positive performance to a managed fund, then that fund adds value for at least one investor.

Ferson and Lin (2014) address the issue whether an investor should buy a positive alpha fund. Assume that markets are incomplete, and clients make optimal consumption/investment decisions to maximize time separable lifetime utility. The client has a stochastic discount factor given by $m_{t+1} = \beta J'(W_{t+1})/U'(C_t)$, where β is the subjective time discount factor, J is the indirect value function, W_{t+1} is the wealth at time $t+1$, and C_t is consumption at time t . The client has access to public information Z_t and the pricing equation holds for the set of N primitive assets that the client trades in.

The client is faced with the opportunity to invest in a fund manager who trades in the same set of N assets but has access to information Ω_t , which also includes Z_t . The fund manager offers a return of R_{pt+1} . If the fund offers a nonzero α_p relative to the client's stochastic discount factor, then the client will adjust their optimal consumption/investment decisions until α_p equals zero. Proposition 1 in Ferson and Lin (2014) derives the optimal weight invested in R_p and shows that the client will buy a positive alpha fund regardless of how the positive alpha is generated.⁴

The analysis in Ferson and Lin (2014) shows that the relevant benchmark to evaluate fund performance is the marginal utility growth of the client. In incomplete markets, investors can have different preferences, and so the marginal utility growth can vary across clients. Ferson (2010, 2013) advocates the use of clientele specific fund performance measures.

There are two main approaches to construct candidate stochastic discount factor models to evaluate fund performance. The first is based on an asset pricing model, of which the most common has been linear factor models.⁵ Linear factor models, such as the capital asset pricing model (CAPM) or arbitrage pricing theory (APT), imply the following stochastic discount factor:

$$m_{t+1} = a + \sum_{k=1}^K b_k f_{kt+1} \quad (7)$$

where a and b_k are the constant and slope coefficients on the k th factor in the stochastic discount factor, f_{kt+1} is the value of the k th factor at time $t+1$, and K is the number of factors in the model. The b_k slope coefficient is linked to the risk premium of the factor (Cochrane (2005)). Equation (7) is an unconditional factor model. A conditional factor model is one where a_t and b_{kt} can vary through time as a function of the information set of investors.⁶

⁴Proposition 2 of an earlier version of the Ferson and Lin (2014) paper shows that if the client has an indirect value function that is quadratic in wealth, then an informed investor will generate a positive alpha.

⁵Cochrane (2005) and Ferson (2019) show the equivalence between the stochastic discount factor approach, mean-variance efficiency, and expected return/beta models.

⁶Hansen and Richard (1987) show that conditional factor models are untestable as the information set of investors is unobservable.

Linear factor models in Eq. (7) can include economic factors or statistical factors but in the main have used portfolio factors such as Fama and French (1993) or Carhart (1997). A partial list of studies that have used stochastic discount factors based on linear factor models to evaluate fund performance include Farnsworth et al. (2002), Li et al. (2016), Chretien and Kammoun (2018) among others. Some studies have also considered using nonlinear stochastic discount factors based on the consumption CAPM such as Ahn et al. (2009), Chretien and Kammoun, and Gredil et al. (2020).

There are two major problems with using asset pricing models to evaluate fund performance. The first problem is that asset pricing models struggle to correctly price the N test assets and so result in inadmissible stochastic discount factors (see Hodrick and Zhang (2001)). The second problem is that using a given asset pricing model assumes that all investors will agree on the value of the fund. Ferson and Lin (2014) examine the impact of investor heterogeneity on fund performance measures based on factor models and find it is economically significant. Chretien and Kammoun (2018) also find that this is a major issue of US equity mutual funds.

The second main approach to evaluate fund performance is to identify admissible stochastic discount factors from a set of N test assets, which requires weaker economic assumptions. Chen and Knez (1996) use the minimum variance LOP and NA admissible stochastic discount factors of Hansen and Jagannathan (1991) to evaluate US mutual funds. LOP and NA admissible stochastic discount factors used by Chen and Knez are given by:

$$m_{\text{LOP},t+1} = R_{t+1}\lambda, \text{ and } m_{\text{NA},t+1} = \max(0, R_{t+1}\lambda) \quad (8)$$

where R_{t+1} is a (T, N) matrix of gross returns of test assets at time $t+1$, and λ is a $(N, 1)$ vector of constants. Under the LOP, $\lambda = E(R'_{t+1}R_{t+1})^{-1}1_N$, where 1_N is a $(N, 1)$ vector of ones.

Ahn et al. (2009a) derive the NA performance bounds for a given fund assigned by all NA admissible stochastic discount factors. The NA bounds are calculated as follows:

$$\begin{aligned} \alpha_{\text{NA},U} &= \text{Max}(m_{t+1}) E(m_{t+1}R_{pt+1}) - 1 \text{ subject to } E(m_{t+1}R_{t+1}) \\ &= 1_N, \text{ and } m_{t+1} \geq 0 \end{aligned} \quad (9)$$

$$\begin{aligned} \alpha_{\text{NA},L} &= \text{Min}(m_{t+1}) E(m_{t+1}R_{pt+1}) - 1 \text{ subject to } E(m_{t+1}R_{t+1}) \\ &= 1_N, \text{ and } m_{t+1} \geq 0 \end{aligned} \quad (10)$$

When $\alpha_{\text{NA},L} > 0$, then all NA admissible stochastic discount factors assign positive performance to the fund. When $\alpha_{\text{NA},U} < 0$, then all NA admissible stochastic discount factors assign negative performance to the fund. When the NA performance bounds straddle zero, then NA admissible stochastic discount factors disagree on the sign of the performance of the fund.

Ahn et al. (2009) interpret the upper NA performance bound of the fund as representing the performance of the fund to investors most favorable to the fund or

the best clientele (BC) of investors. This provides one way to create clientele specific performance measures, as advocated by Ferson (2010, 2013). The width of the bounds provides an indication of the magnitude of investor disagreement over the performance of a given fund. The NA performance bounds can often be extremely wide. Ahn et al. also propose using the bounds as a diagnostic test of performance measures based on asset pricing models. If the fund performance based on asset pricing models does not lie within the NA bounds, then the performance is inadmissible.

Chretien and Kammoun (2017) extend the analysis of Ahn et al. (2009) and develop a BC performance measure based on the good-deal option pricing bounds of Cochrane and Saa-Requejo (2000). Cochrane and Saa-Requejo place an upper bound on the maximum Sharpe (1966) performance available to investors, which puts an upper bound on $E(m^2)$ (and $\sigma(m)$) of admissible stochastic discount factors. A Sharpe performance higher than the maximum is known as good deal, which investors would exploit. The BC performance measure is given by:

$$\alpha_{BC} = \max (m_{t+1}) E(m_{t+1} R_{pt+1}) - 1 \quad (11)$$

subject to $E(m_{t+1} R_{t+1}) = 1_N$, and $E(m_{t+1}^2) \leq (1 + h^2)/R_F^2$ where h is the maximum Sharpe ratio, and R_F is the gross risk-free return. The lower good-deal bound solves the $\min (m) (m_{t+1}) E(m_{t+1} R_{pt+1}) - 1$. Under these restrictions, Cochrane and Saa-Requejo show that the good-deal bounds can be solved analytically.⁷

Chretien and Kammoun (2017) show that the difference between $\alpha_{BC} - \alpha_{LOP} = vE(w_{t+1}^2)$, where v is the disagreement measure and is given in Eq. (15) of their paper, and w_{t+1} is the replication error of fund. In order to estimate the BC performance of a fund, we need to specify an upper bound on the maximum Sharpe performance. Cochrane and Saa-Requejo (2000) set the maximum Sharpe performance to be twice the Sharpe performance of the S&P 500 index.⁸ In estimating the BC performance, the maximum Sharpe performance has to be set higher than the maximum Sharpe performance (h^*) that is available in the test assets. Chretien and Kammoun set $h = h^* + \text{Mult} \cdot h_{MKT}$, where h_{MKT} is the Sharpe performance of the market index, and Mult is the multiple chosen. In their empirical application, they focus on setting Mult equal to 0.5 or 1.

Chretien and Kammoun (2018) use the BC performance measure as a diagnostic tool of performance measures. Define α_y as the performance of a fund using a candidate stochastic discount factor model such as linear factor models. If $\alpha_{BC} = \alpha_y$, then the performance using the candidate model gives the same performance as the performance from the BC of investors. Chretien and Kammoun consider two

⁷Cochrane and Saa-Requejo (2000) also impose the NA restriction when estimating good-deal bounds, but this requires the use of numerical methods. Chretien and Kammoun (2017) find a tiny proportion of BC admissible stochastic discount factors take on negative values, and so the NA restriction is likely to have little impact on the BC performance.

⁸Related applications include MacKinlay (1995), Huang and Zhou (2017), and Barillas and Shanken (2018) among others.

alternative hypotheses. First, if $\alpha_{BC} < \alpha_\gamma$, then the performance using the candidate model is inadmissible. Second, if $\alpha_{BC} > \alpha_\gamma$, then the performance using the candidate model suffers from a “misrepresentation” problem, in that the candidate model misrepresents the performance of the BC of investors.

The empirical evidence in Ferson and Lin (2014), and Chretien and Kammoun (2017, 2018, 2019), among others, suggest that investor heterogeneity is an important issue in evaluating mutual fund performance. A recent study by Almeida et al. (2020) provides an alternative approach to incorporate investor heterogeneity in fund performance based on minimum discrepancy (MD) admissible stochastic discount factors. The MD approach allows for nonlinear functions of the primitive assets, and takes account of higher moments beyond mean and variance, and so is relevant for evaluating the performance of managed funds that engage in dynamic trading, or invest in derivatives, or use leverage such as hedge funds. The MD approach allows investors to have different preferences for higher moments.

Almeida et al. (2020) build on the earlier studies of Almeida and Garcia (2012, 2017), and use the Cressie and Read (1984) family of discrepancies. The MD admissible stochastic discount factors are estimated to solve:

$$\text{Min } (m_{t+1}) E[m_{t+1}^{\gamma+1} / (\gamma(\gamma+1))] \quad (12)$$

subject to $E(m_{t+1}) = 1$,⁹ $E(m_{t+1}(R_{t+1}-1)) = 0_N$, and $m_{t+1} > 0$, where γ is the selected power function, and 0_N is a $(N,1)$ vector of zeros. Given the MD admissible stochastic discount factor, we can then estimate the performance of the fund. Almeida et al. solve the dual problem related to the primal problem in Eq. (12). Underlying the solution to the MD admissible stochastic discount factor is a vector of Lagrange Multipliers (λ_N), which is a $(N,1)$ vector. Almeida et al. show that by varying γ , we can capture different investor preferences to higher moments such as skewness and kurtosis, and so provides a way to examine the impact of investor heterogeneity on fund performance.

13.3 Estimation Issues

The estimation of fund performance, within the stochastic discount factor framework, is most often conducted using Generalized Method of Moments¹⁰ (GMM) (Hansen (1982)). This section will focus on estimation issues using linear factor models as the candidate stochastic discount factor models as these models are the most common in the evaluation of fund performance. This section will focus on return-based performance measures, where the researcher does not have access to the

⁹The analysis can be adapted for alternative values of $E(m)$.

¹⁰See Hall (2005) for an excellent textbook treatment of GMM.

underlying portfolio weights of the funds.¹¹ GMM estimation along with other empirical methods have been used to evaluate fund performance using admissible stochastic discount factor models covered in the prior section.

When using linear factor models, there are different approaches that can be adopted. The first approach is to use a two-step method. First, the coefficients in the linear factor model are estimated to calculate the fitted stochastic discount factor values of the model. Second, given the estimated stochastic discount factor the performance of the fund can be estimated by the sample moment of eq. (5). The second approach is to estimate the stochastic discount factor coefficients, and fund performance jointly. Farnsworth et al. (2002) advocate this approach as it is more efficient than a two-step approach as it takes account of the sampling error in estimating the coefficients in the stochastic discount factor, when evaluating fund performance. The two-step approach can also do this if we use the delta method to derive the standard errors of the fund performance measures, which is function of the covariance matrix of the estimated stochastic discount factor coefficients.

Define the following set of residuals for a given linear factor model y_{t+1} using the excess returns of the fund (r_{pt+1}):

$$u_{it+1} = r_{it+1}y_{t+1} - 0 \quad \text{for } i = 1, \dots, N \quad (13)$$

$$u_{pt+1} = \alpha_p - y_{t+1}r_{pt+1}$$

where α_p is the unconditional performance of the fund. The sample mean of the residuals ($E(u_{it+1})$, and $E(u_{pt+1})$) are the $(N+1)$ vector of moment conditions (g) in GMM estimation. The first N moment conditions identify the $K+1$ parameters in the linear factor model¹² y_{t+1} and are also the pricing errors of the N test assets. The last moment condition identifies the performance of the fund. There are $N+1$ moment conditions and $K+2$ parameters in the system of equations in (13). Under the null hypothesis of no abnormal performance, α_p should be equal to zero.

GMM estimates the $K+2$ parameters in Eqs. (13) to minimize the quadratic form $g'Wg$, where W is an $(N+1, N+1)$ arbitrary weighting matrix. Hansen (1982) shows that the estimates have an asymptotic normal distribution for any arbitrary weighting matrix. Hansen also shows that the most efficient weighting matrix is S^{-1} , where S is the spectral density matrix of the moment conditions. The advantage of the GMM approach is that it is valid under general distributional assumptions and we can incorporate the effects of serial correlation and heteroskedasticity.

When evaluating the performance of a large number of funds, the number of moment conditions increases sharply. However, Farnsworth et al. (2002) show that

¹¹Ferson and Mo (2016) extend the stochastic discount factor approach to evaluate fund performance when the portfolio weights of the fund are available. They also provide a performance decomposition for a given fund into stock selection ability, factor timing ability, and volatility timing ability. See also Ferson and Wang (2021).

¹²It is straightforward to extend the analysis to allow for conditional factor models along the lines of Cochrane (1996).

the estimated performance and standard error of the fund is invariant to the number of funds used in the estimation. This result implies there are no biases in the performance (or standard error) for a given fund by estimating for one fund at a time. Ferson et al. (2006) generalize this result to conditional time-varying performance measures.

Additional moment conditions can be added to the system of equations in (13), to incorporate additional restrictions implied by stochastic discount factor models. Dahlquist and Soderlind (1999) and Farnsworth et al. (2002) show that it is important to add a moment condition for the gross risk-free return. The expected value of the stochastic discount factor should be just below one (see Cochrane (2005)). By including this moment condition, the expected value of the stochastic discount factor model has a more sensible value. Farnsworth et al. also point out that for linear factor models of the stochastic discount factor, where the factors are portfolio returns, it can be useful to impose the restriction that the model correctly prices the factors.

Three issues arise using GMM to estimate the unconditional performance of the fund. First, the researcher must choose the set of N test assets. This set should capture the investment opportunity set that investors can trade in. The number of test assets should also be small. Cochrane (2005) recommends that the number of moment conditions should be at most 1/10th of the number of time-series observations because the estimate of S can become unstable and near singular when the number of moment conditions is too high.

Different sets of test assets have been used in the literature. Chen and Knez (1996), Dahlquist and Soderlind (1999), and Ahn et al. (2009) all use industry portfolios in their set of primitive assets. Dahlquist and Soderlind also add a short-term Treasury Bill to the set of test assets. Farnsworth et al. (2002) use two bond portfolios, one-month Treasury Bill, and six stock portfolios that capture small cap/large cap, value/growth, and momentum/contrarian investment strategies. Chretien and Kammoun (2017) also use six size/book-to market (BM) portfolios as one set of test assets used in their study. One question that arises is whether fund performance is sensitive to the choice of the test assets. Ahn et al. (2009b) explore this issue in an asset pricing context and propose a novel approach to form the set of test assets.

Second, what weighting matrix should be used in estimating the system of equations in (13), when the number of parameters is less than the number of moment conditions? A major problem in using the optimal weighting matrix S^{-1} is that in small sample sizes, the optimal weighting matrix can perform poorly (Lettau and Ludvigson (2001)). The optimal weighting matrix can suffer from two other problems whenever we want to consider how well different factor models of y_{t+1} price the N test assets. First, the optimal weighting matrix is different across each model and so cannot be used to compare the pricing performance of different models. Using the optimal weighting matrix, improvements in model performance can either come from lower pricing errors and/or a more volatile weighting matrix. Second, the optimal weighting matrix evaluates the ability of stochastic discount factor models to price the test assets in terms of how well the model prices portfolios of large long and short positions (Cochrane (2005)). There is also the issue of whether we use the

two-step approach of optimal GMM as in Hansen and Singleton (1982), or iterated optimal GMM (Ferson and Foerster (1994)), or continuous-updating optimal GMM (Hansen et al. (1996)).

The two most popular alternative weighting matrixes that can be used to evaluate fund performance are ones proposed in the asset pricing literature. Cochrane (1996) advocates the use of the identity weighting matrix and Hansen and Jagannathan (1997) advocate the use of the inverse of the second moment matrix of asset payoffs. Since the same weighting matrix is used across models, we can evaluate how well different models price the test assets. The identity weighting matrix places an equal weight on each moment condition. In terms of evaluating asset pricing models, the identity matrix estimates the parameters to minimize the sum of squared pricing errors. This approach is most useful whenever the researcher wants to examine how well models price a given set of assets rather than complex long/short portfolios of assets. However, the use of the identity weighting matrix can lead to more volatile estimates of the parameters (Hodrick and Zhang (2001)).

The use of the Hansen and Jagannathan (1997) weighting matrix in the GMM minimization $g'Wg$ is equal to the squared Hansen and Jagannathan distance measure under the LOP assumption (First HJD). The First HJD measure captures the smallest distance between a given candidate model of the stochastic discount factor and the true set of discount factors that price the primitive assets. The First HJD measure is also the most mispriced portfolio of the test assets with unit norm. Asset pricing models that are more able to price the test assets should have lower First HJD measures^{13 14}.

Third, the standard errors and statistical tests in the system of equations in (13) assume that the factor model is correctly specified (Hall (2005)), i.e., the pricing errors of the test assets are zero. Hall and Inoue (2003) derive the asymptotic distribution of some GMM estimators allowing for potential model misspecification. A series of papers derive the asymptotic distribution of the stochastic discount factor coefficients of linear factor models using different GMM weighting matrixes, including Kan and Robotti (2009), Gospodinov et al. (2013, 2014, 2018), among others, allowing for potential model misspecification. The impact of allowing for potential model misspecification is that it tends to increase the standard errors of the stochastic discount factor coefficients, especially for economic factors. It also increases the sampling variation in model specification tests. An interesting extension to this literature would be to examine what impact of allowing for potential model misspecification tests has on the standard errors of fund performance.

One solution to these issues over the choice of test assets, the choice of GMM weighting matrix, and dealing with potential model misspecification is to estimate

¹³ Hansen and Jagannathan (1997) propose a NA distance measure (Second HJD) which captures the smallest distance between the candidate stochastic discount factor model and the set of NA admissible stochastic discount factors.

¹⁴ Model comparison tests have been developed using the HJD measures including Kan and Robotti (2009), Li et al. (2010), and Gospodinov et al. (2013, 2016) among others.

the system of equations in (13) setting $N + 1 = K + 2$. This approach can be done for linear factor models, by including the gross Treasury Bill return and the excess returns of the factors as the set of N test assets. The stochastic discount factor coefficients are estimated so that the gross Treasury Bill returns and the excess factor returns are correctly priced. This approach has been used by De Souza and Lynch (2012), and Ferson and Mo (2016) among others. The downside of this approach is of course we ignore the pricing problems the models might have for other sets of test assets.

The system of equations in (13) can be extended to allow for conditional fund performance. We can estimate time-varying conditional performance by assuming that the performance of the fund is a linear function of lagged common information variables as in Dahlquist and Soderlind (1999), and De Souza and Lynch (2012) among others. An alternative approach to the linear functional form is followed by Ferson et al. (2006) who use a small number of conditioning dummy variables based on a lagged information variable that capture different states of the economy, and then estimate fund performance within each state.

13.4 Conclusions

The stochastic discount factor approach provides a unified framework to evaluate fund performance. It can be applied using a wide array of models and can be used to evaluate the performance of any type of managed fund. No doubt the stochastic discount factor approach will continue to be widely used to evaluate the performance of different types of funds such as pension funds, hedge funds, and private equity funds among others.

References

- Ahn, D.H., H.H. Cao, and S. Chretien. 2009a. Portfolio performance measurement: A no arbitrage bounds approach. *European Financial Management* 15: 298–339.
- Ahn, D.H., J. Conrad, and R.F. Dittmar. 2009b. Basis assets. *Review of Financial Studies* 22: 55133–55174.
- Almeida, C., K. Ardison, and R. Garcia. 2020. Nonparametric assessment of hedge fund performance. *Journal of Econometrics* 214: 349–378.
- Almeida, C., and R. Garcia. 2012. Assessing misspecified asset pricing models with empirical likelihood estimators. *Journal of Econometrics* 170: 519–537.
- . 2017. Economic implications of nonlinear pricing kernels. *Management Science* 63: 3361–3380.
- Aragon, G., and W.E. Ferson. 2008. Portfolio performance evaluation. *Foundations and Trends in Finance* 2: 83–190.
- Barillas, F., and J. Shanken. 2018. Comparing asset pricing models. *Journal of Finance* 73: 715–754.
- Carhart, M.M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52: 57–82.
- Chen, Z., and P.J. Knez. 1996. Portfolio performance measurement: Theory and applications. *Review of Financial Studies* 9: 511–555.

- Chan, L.K.C., S.G. Dimmock, and J. Lakonishok. 2009. Benchmarking money manager performance: Issues and evidence. *Review of Financial Studies* 22: 4553–4599.
- Chretien, S., and M. Kammoun. 2017. Mutual fund performance evaluation and best clienteles. *Journal of Financial and Quantitative Analysis* 52: 1577–1604.
- . 2018. Representative investors versus best clienteles: Performance evaluation disagreement in mutual funds, Working Paper, University of Laval.
- . 2019. Mutual fund styles and clientele-specific performance evaluation. *International Journal of Economics and Finance* 11: 89–116.
- Cochrane, J.H. 1996. A cross-sectional test of an investment based asset pricing model. *Journal of Political Economy* 104: 572–621.
- . 2005. *Asset pricing*: Revised edition. Princeton: Princeton University Press.
- Cochrane, J.H., and J. Saa-Requejo. 2000. Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *Journal of Political Economy* 108: 79–119.
- Cressie, N., and T.R.C. Read. 1984. Multinomial goodness-of-fit tests. *Journal of the Royal Statistics Society Series B* 46: 440–464.
- Dahlquist, M., and P. Soderlind. 1999. Evaluating portfolio performance with stochastic discount factors. *Journal of Business* 72: 347–383.
- De Souza, A., and A.W. Lynch. 2012. Does mutual fund performance vary over the business cycle?, Working Paper, New York University.
- Fama, E.F., and K.R. French. 1993. Common risk factors in the returns on bonds and stocks. *Journal of Financial Economics* 33: 3–53.
- Farnsworth, H., W.E. Ferson, D. Jackson, and S. Todd. 2002. Performance evaluation with stochastic discount factors. *Journal of Business* 75: 473–505.
- Ferson, W.E. 2010. Investment performance evaluation. *Annual Review of Financial Economics* 2: 207–234.
- . 2013. Ruminations on investment performance measurement. *European Financial Management* 19: 4–13.
- . 2019. *Empirical asset pricing: Models and methods*. MIT Press.
- Ferson, W.E., and S.R. Foerster. 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. *Journal of Financial Economics* 36: 29–55.
- Ferson, W.E., T. Henry, and D. Kisgen. 2006. Evaluating government bond fund performance with stochastic discount factors. *Review of Financial Studies* 19: 423–456.
- Ferson, W.E., and K. Khang. 2002. Conditional performance measurement using portfolio weights: Evidence for pension funds. *Journal of Financial Economics* 65: 249–282.
- Ferson, W.E., and J. Lin. 2014. Alpha and performance measurement: The effects of investor disagreement and heterogeneity. *Journal of Finance* 69: 1565–1596.
- Ferson, W.E., and H. Mo. 2016. Performance measurement with selectivity, market and volatility timing. *Journal of Financial Economics* 121: 93–110.
- Ferson, W.E., and R.W. Schadt. 1996. Measuring fund strategy and performance in changing economic conditions. *Journal of Finance* 51: 425–461.
- Ferson, W.E., and J.L. Wang. 2021. A panel regression approach to holdings-based fund performance measures. *Review of Asset Pricing Studies* (forthcoming).
- Glosten, L., and R. Jagannathan. 1994. A contingent claim approach to performance evaluation. *Journal of Empirical Finance* 1: 133–160.
- Gospodinov, N., R. Kan, and C. Robotti. 2013. Chi-squared tests for evaluation and comparison of asset pricing models. *Journal of Econometrics* 173: 108–125.
- . 2014. Misspecification-robust inference in linear asset pricing models with irrelevant risk factors. *Review of Financial Studies* 27: 2139–2170.
- . 2016. On the properties of the constrained Hansen-Jagannathan distance. *Journal of Empirical Finance* 36: 121–150.
- . 2018. Asymptotic variance approximations for invariant estimators in uncertain asset-pricing models. *Econometric Reviews* 37: 695–718.

- Gredil, O., M. Sorensen, and W. Waller. 2020. Evaluating private equity performance using stochastic discount factors, Working Paper, Dartmouth College.
- Hall, A.R. 2005. *Generalized method of moments*. Oxford University Press.
- Hall, A.R., and A. Inoue. 2003. The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics* 114: 361–394.
- Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.
- Hansen, L.P., J. Heaton, and A. Yaron. 1996. Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics* 14: 262–280.
- Hansen, L.P., and R. Jagannathan. 1997. Assessing specification errors in stochastic discount factor models. *Journal of Finance* 52: 591–607.
- Hansen, L.P., and S.F.R. Richard. 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* 55: 587–613.
- Hansen, L.P., and K.J. Singleton. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50: 1269–1268.
- Harrison, M., and D. Kreps. 1979. Martingales and arbitrage in multi-period securities markets. *Journal of Economic Theory* 20: 381–408.
- Hodrick, R., and X. Zhang. 2001. Evaluating the specification errors of asset pricing models. *Journal of Financial Economics* 62: 327–376.
- Huang, D., and G. Zhou. 2017. Upper bounds on return predictability. *Journal of Financial and Quantitative Analysis* 52: 401–425.
- Kan, R., and C. Robotti. 2009. Model comparison using the Hansen-Jagannathan distance. *Review of Financial Studies* 22: 3449–3490.
- Korteweg, A., and S. Nagel. 2016. Risk-adjusting the returns to venture capital. *Journal of Finance* 71: 1437–1470.
- Lehmann, B.N., and D.M. Modest. 1987. Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance* 42: 233–265.
- Lettau, M., and S. Ludvigson. 2001. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109: 1238–1287.
- Li, H., Y. Xu, and X. Zhang. 2010. Evaluating asset pricing models using the second Hansen-Jagannathan distance. *Journal of Financial Economics* 97: 279–301.
- . 2016. Hedge fund performance evaluation under the stochastic discount factor framework. *Journal of Financial and Quantitative Analysis* 51: 231–257.
- MacKinlay, A.C. 1995. Multifactor models do not explain deviations from the capital asset pricing model. *Journal of Financial Economics* 38: 3–28.
- Roll, R. 1978. Ambiguity when performance is measured by the securities market line. *Journal of Finance* 33: 1051–1069.
- Ross, S.A. 1978. A simple approach to the valuation of risky streams. *Journal of Business* 51: 153–475.
- Sharpe, W.F. 1966. Mutual fund performance. *Journal of Business* 39: 119–138.

Duration Concepts, Analysis, and Applications

14

Zvika Afik, Iraj Fooladi, Gady Jacoby, and Gordon Roberts

Abstract

We discuss duration and its development, placing particular emphasis on various applications. The survey begins by introducing duration and showing how traders and portfolio managers use this measure in speculative and hedging strategies. We then turn to convexity, a complication arising from relaxing the linearity assumption in duration. Next, we present immunization – a hedging strategy based on duration – and then examine stochastic process risk, foreign-exchange risk, and duration extensions that address these risks. We also examine the track record of duration and how the measure applies to financial futures. The discussion then turns to macrohedging the entire balance sheet of a financial institution. We develop a theoretical framework for duration gaps and apply it, in turn, to banks, life insurance companies, and defined benefit pension plans.

For *The Encyclopedia of Finance*, C.F. Lee, editor, Third Edition, Kluwer Academic Publishers.

Gordon Roberts: deceased.

Z. Afik

Management of Service Organizations Department, Hadassah Academic College, Jerusalem, Israel
e-mail: zvikaafik@hac.ac.il

I. Fooladi (✉)

Rowe School of Business, Dalhousie University, Halifax, NS, Canada
e-mail: iraj.fooladi@dal.ca

G. Jacoby

I.H. Asper School of Business, The University of Manitoba, Winnipeg, MB, Canada
e-mail: Gady.jacoby@umanitoba.ca

G. Roberts

Schulich School of Business, York University, Toronto, ON, Canada

Keywords

Duration · Fixed income securities · Immunization · Hedging interest rate risk · Macrohedging · Bond price volatility · Stochastic process risk · Exchange rate risk · Financial institution management · Pension funds · Insurance companies · Banks

14.1 Introduction

Duration Analysis is the key to understanding the returns on fixed income securities. Duration is also central to measuring risk exposures in fixed income positions.

The concept of duration was first developed by Macaulay (1938). Thereafter, it was occasionally used in some applications by economists (Hicks 1939; Samuelson 1945) and actuaries (Redington 1952). However, by and large, this concept remained dormant until 1971 when Fisher and Weil illustrated that duration could be used to design a bond portfolio that is immunized against interest rate risk. Today, the concepts in this chapter are widely used in financial markets.

Bierwag and Fooladi (2006) survey the development of the theory on historical context. Here we discuss duration and its development, placing particular emphasis on various applications. The survey begins by introducing duration and showing how traders and portfolio managers use this measure in speculative and hedging strategies. We then turn to convexity, a complication arising from relaxing the linearity assumption in duration. We also discuss how Eq. (4) below could be used to calculate Value at Risk. Next, we present immunization – a hedging strategy based on duration – and introduce the concept of contingent immunization as a mixed active and passive bond-portfolio management strategy. Then we discuss the risk arising from exposure to currency exchange volatility of foreign bonds, presenting a duration model addressing risks associated with both interest rate and foreign exchange rate. The chapter goes on to examine stochastic process risk and duration extensions that address it. We examine the track record of duration and how the measure applies to financial futures and then turn to the duration of bonds subject to default risk and to macrohedging applied to the entire balance sheet of a financial institution. We develop a theoretical framework for duration gaps and apply it, in turn, to banks, life insurance companies, and defined benefit pension plans. We then show how it would be possible to simultaneously protect net worth against changes in inflation and real interest rate. Finally, we briefly discuss other applications of duration gap.

14.2 Duration Formulation and Interpretation

Recognizing that term-to-maturity of a bond was not an appropriate measure of its actual life, Macaulay (1938) invented the concept of duration as the true measure of a bond's "longness" and applied the concept to asset/liability management of life

insurance companies. Thus, duration represents a measure of the time dimension of a bond or other fixed income security. The formula calculates a weighted average of the time horizons at which cash flows from a fixed income security are received. Each time horizon's weight is the percentage of the total present value of the bond (bond price) paid at that time. These weights add up to one. Macaulay duration, given in Eq. (1), uses the bond's yield to maturity to calculate the present values.

$$\text{Duration} = D = \frac{\sum_{t=1}^N \frac{tC(t)}{(1+y)^t}}{P_0} = \sum_{t=1}^N tW(t) \quad (1)$$

where

$C(t)$ = cash flow received at time t ,

$W(t)$ = weight attached to time t cash flow, $\sum_{t=1}^N W(t) = 1$, and

y = yield-to-maturity

P_0 = current price of the bond, $P_0 = \sum_{t=1}^N \frac{C(t)}{(1+y)^t}$

A bond's duration increases with maturity, but is shorter than maturity unless the bond is a zero-coupon bond (in which case it is equal to maturity). The coupon rate also affects duration because a bond with a higher coupon rate pays a greater percentage of its present value prior to maturity. Such a bond has greater weights on coupon payments and hence a shorter duration.

Using yield to maturity to obtain duration implies that interest rates are the same for all maturities (a flat term structure). Equation (2) provides an expression for Fisher and Weil's (1971) reformulation of duration using a more general (non-flat) term structure. They further show that duration can be used to immunize a portfolio of fixed income securities.

$$D = \frac{\sum_{t=1}^N \frac{tC(t)}{(1+r_t)^t}}{P_0} = \sum_{t=1}^N tW(t) \quad (2)$$

where, r_t = discount rate for cashflows received at time t .

Their development marks the beginning of a broader application to active and passive fixed income investment strategies, which came in the 1970s as managers looked for new tools to address the sharply increased volatility of interest rates.¹ In general, duration has two practical properties.

¹Bierwag (1977, 1987a) provide excellent reviews of duration analysis.

- Duration represents the *elasticity* of a bond's price with respect to the discount factor $(1 + y)^{-1}$. This was first developed by Hicks (1939). This property has applications for active bond portfolio strategies and evaluating *Value at Risk*.
- When duration is maintained equal to the time remaining in an investment planning horizon, promised portfolio return is immunized.

14.3 Duration and Price Volatility

In analyzing a series of cash flows, Hicks (1939) calculates the elasticity of the series with respect to the discount factor, which results in rederiving the Macaulay duration. Noting that this elasticity is defined in terms of time, he calls it “average period” and shows that the relative price of two series of cash flows with the same average period is unaffected by changes in interest rates. Hicks' work brings our attention to a key mathematical property of duration. *The price elasticity of a bond in response to a small change in its yield to maturity is proportional to duration.* Following the essence of work by Hopewell and Kaufman (1973), we can approximate the elasticity as:

$$\text{Duration} = D = -\frac{dP}{dr} \frac{(1+r)}{P} \cong -\left[\frac{\Delta P/P}{\Delta r/(1+r)} \right] \quad (3)$$

where P denotes the price of the bond and r denotes the market yield. Rearranging terms, we obtain:

$$\Delta P \cong -D[\Delta r/(1+r)]P \quad (4)$$

When interest rates are continuously compounded, Eq. (4) turns to:

$$\Delta P \cong -D[\Delta r]P$$

This means that if interest rates fall (rise) slightly, the price increases (decreases) in different bonds are proportional to duration.

The intuition here is straightforward: if a bond has a longer duration because a greater portion of its cash flows are being deferred further into the future, then a change in the discount factor has a greater effect on its price. Note again that, here, we are using yield instead of term structure and, thus, strictly speaking, assuming a flat term structure.

The link between bond duration and price volatility has important practical applications in trading, portfolio management, and managing risk positions. For the trader taking a view on the movement of market yields, duration provides a measure of volatility or potential gains. Other things equal, the trader will seek maximum returns to a rate anticipation strategy by taking long or short positions in high duration bonds. For derivative strategies, price sensitivity for options and futures contracts on bonds also depends on duration. In contrast with traders, bond

portfolio managers have longer horizons. They remain invested in bonds but lengthen or shorten portfolio average duration depending on their forecast for rates.

14.4 Convexity: A Duration Complication

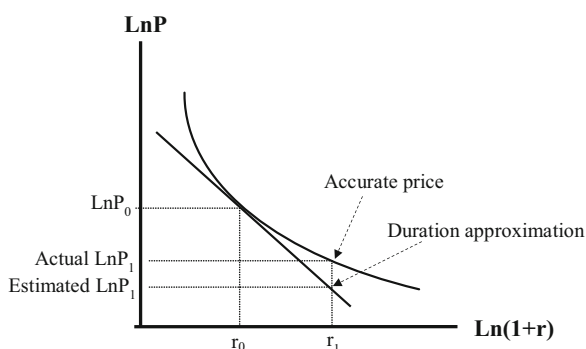
Equation (4) is accurate for small shifts in yields. In practice, more dramatic shifts in rates sometimes occur. For example, in response to the financial crisis of 2007–2008, central banks around the world dramatically lower interest rates, raising the prices and the interest rate sensitivity of all bonds. In cases where interest rate changes involve such large shifts, the price changes predicted from the duration formula are only approximations. The cause of the divergence is convexity. To understand this argument better, note that the duration derived in Eq. (3) can be rewritten as:

$$D \cong - \left[\frac{\Delta P/P}{\Delta(1+r)/(1+r)} \right] = - \frac{\Delta \ln P}{\Delta \ln(1+r)} \quad (5)$$

However, the true relationship between $\ln P$ and $\ln(1+r)$ is represented by a convex function. Duration is the absolute value of the slope of a line, which is tangent to the curve representing this true relationship. A curve may be approximated by a tangent line only near the point of tangency.

Figure 1 illustrates convexity plotting the *natural log of bond price* on the y-axis and the *natural log of 1 plus interest rate* on the x-axis. The absolute value of the slope of the straight line that is tangent to the actual relationship between price and interest rate at the present interest rate represents the duration. Figure 1 shows that the duration model gives very accurate approximations of percentage price changes for small shifts in yields. As the yield shifts become larger, the approximation becomes less accurate and the error increases. Duration overestimates the price decline resulting from an interest rate hike and underestimates the price increase caused by a decline in yields. This error is caused by the convexity of the curve representing the true relationship.

Fig. 1 Actual versus duration estimate for changes in the bond price



Thus, convexity (sometimes called positive convexity) is “good news” for an investor with a long position: when rates fall, the true price gain (along the curve) is greater than predicted by the duration line. On the other hand, when rates rise, the true percentage loss is smaller than predicted by the duration line.

Note that the linear price-change relationship ignores the impact of interest rate changes on duration. In reality, duration is a function of the level of rates because the weights in the duration formula all depend on bond yield. Duration falls (rises) when rates rise (fall) because a higher discount rate lowers the weights for cash flows far into the future. These changes in duration cause the actual price-change curve to lie above the tangent line in Fig. 1. The positive convexity described here characterizes all fixed income securities that do not have embedded options such as call or put features on bonds or prepayment or lock-in features in mortgages.

Embedded options can cause negative convexity. This property is potentially dangerous as it reverses the “good news” feature of positive convexity as actual price falls below the level predicted by duration alone.

Researchers have proposed to include convexity when is required for macro-hedging (Fooladi and Roberts 2004) and for immunization (Zhu et al. 2018).

14.5 Duration and Value at Risk

Financial institutions face market risk as a result of the actions of the trader and the portfolio manager. Market risk occurs when rates move opposite to the forecast on which an active strategy is based. For example, a trader may go short and will lose money if rates fall. In contrast, a portfolio manager at the same financial institution may take a long position with higher duration and will face losses if rates rise.

Value at Risk methodology makes use of Eq. (4) to calculate the institution’s loss exposure.² For example, suppose that the net position of the trader and the portfolio manager is \$50 million ($P = \50 million) in a portfolio with a duration of 5 years. Suppose further that the worst-case scenario is that rates, currently at 6%, jump by 50 basis points in one day ($\Delta r = 0.005$). The risk management professional calculates the maximum loss or Value at Risk as

$$dP = -5 [0.005 / (1 + .06)] 50 \text{ million} = -1.179 \text{ million}$$

If this maximum loss falls within the institution’s guidelines, the trader and the portfolio manager may not take any action. If, however, the risk is excessive, the treasury professional will examine strategies to hedge the interest rate risk faced by the institution. This leads to the role of duration in hedging.

²For more on market risk see J.P. Morgan (1994) and Cornett and Saunders (2008), Chap. 9.

14.6 Duration and Immunization

Duration hedging or immunization draws on a second key mathematical property. *By maintaining portfolio duration equal to the amount of time remaining in a planning horizon, the investment manager can immunize locking in the original promised return on the portfolio.* Note that immunization seeks to tie the promised return, not to beat it. Because it requires no view of future interest rates, immunization is a passive strategy. It may be particularly attractive when interest rates are volatile.³

Early versions of immunization theory were offered by Samuelson (1945) and Redington (1952). Fisher and Weil (1971) point out that the flat term structure assumption made by Redington and implied in Macaulay duration is unrealistic. They assume a more general (non-flat) term structure of continuously compounded interest rates and a stochastic process for interest rates that is consistent with an additive shift, and prove that *a bond portfolio is immune to interest rate shifts, if its duration is maintained equal to the investor's remaining holding horizon.*

The intuition behind immunization is clearly explained by Bierwag (1987a, Chap. 4). For investors with a fixed planning period, the return realized on their portfolio of fixed income securities could differ from the return they expected at the time of investment as a result of interest rate shifts. The realized rate of return has two components: interest accumulated from reinvestment of coupon income and the capital gain or loss at the end of the planning period. The two components impact the realized rate of return in opposite directions and do not necessarily cancel one another. Which component dominates depends on the relationship between portfolio duration and the length of the planning horizon. When the portfolio duration is longer than the planning period, capital gains or losses will dominate the effect of reinvestment return. This means that the realized return will be less (greater) than the promised return if rates rise (fall). If the portfolio duration is shorter than the planning period, the effect of reinvestment return will dominate the effect of capital gains or losses. In this case, the realized return will be less (greater) than the promised return if rates fall (rise). Finally, when the portfolio duration is exactly equal to the length of the planning period, the portfolio is immunized and the realized return will never fall below the promised rate of return.⁴

Zero-coupon bonds and duration matching with coupon bonds are two ways of immunizing interest rate risk. Duration matching effectively creates synthetic zero-coupon bonds. Equating duration to the planned investment horizon can easily be achieved with a two-bond portfolio. The duration of such a portfolio is equal to the weighted average of the durations of the two bonds that form the portfolio as shown in Eq. (6).

³This discussion of immunization begins by assuming default and option free securities in order to separate interest rate risk from other risks.

⁴Bierwag and Kaufman (1977) maintain that, for the duration matched portfolios, these two effects (unexpected gains and losses resulting from interest rate shifts) cancel each other out, unless the stochastic process is not consistent with the equilibrium conditions, in which case the unexpected gain will be greater than the unexpected loss.

$$D_P = W_1 D_1 + W_2 D_2 \quad (6)$$

where $W_2 = 1 - W_1$. Setting the right-hand side of Eq. (6) equal to the investment horizon, this problem is reduced to solving one equation with one unknown.

The preceding argument is consistent with the view presented in Bierwag and Khang (1979) that immunization strategy is a maxmin strategy: it maximizes the minimum return that can be obtained from a bond portfolio. Prisman (1986) broadens this view by examining the relationship between a duration strategy, an immunization strategy, and a maxmin strategy. He concludes that, for a duration strategy to be able to maximize the lower bound to the terminal value of the bond portfolio, there must be constraints on the bonds to be included.

Ortobelli et al. (2018) suggest a two-step model for bond portfolio management, which they call a “flexible immunization approach.” In essence, under their model’s first step, a modified immunized portfolio is formed and then, under the second step, portfolio performance is enhanced by maximizing its Sharpe and Rachev ratios. Using 2002–2012 U.S. fixed income market data, they test their approach and report an improved performance, particularly during the 2007–2012 period, for which they demonstrate that their approach would have resulted in obtaining lower-risk/higher-return portfolios.

Our discussion on immunization so far includes deterministic interest rate risk measures such as Macaulay and Fisher–Weil durations. Oliveira, Nunes, and Malcato (2014) compare the efficiency of deterministic and stochastic interest rate risk measures. They report that stochastic risk measures provide better immunization strategies than using deterministic measures such as Fisher–Weil duration.

14.7 Contingent Immunization

Since duration is used in both active and passive bond portfolio management, it can also be used for a middle-of-the-road approach. Here fund managers strive to obtain returns in excess of what is possible by immunization; at the same time, they try to limit possible loss from incorrect anticipation of interest rate changes. In this approach, called contingent immunization, the investor sets a minimum acceptable holding period return (HPR) below the promised rate and then follows an active strategy to enhance the HPR beyond the promised return. The investor continues with the active strategy unless, as a result of errors in forecasting, the value of the portfolio shrinks to the point where any further decline will result in an HPR below the minimum limit for the return. At this point, the investor changes from an active to an immunizing strategy.⁵ Diaz et al. (2009) study the performance of contingent-immunization portfolios using mean daily actual transactions in the Spanish public

⁵For more details on contingent immunization, see Bierwag (1987a) and Leibowitz and Weinberger (1981, 1982, 1983).

debt market for bonds, bills, and repos for the period January 4, 1993 to January 3, 2003. They follow strategies that mimic actual contingent immunization strategies and report that, despite its simplicity, “contingent immunization provides excellent performance.” Their performance is at the midpoint between active portfolio and pure immunization strategies.

14.8 Duration of Foreign Bonds

Many foreign-bond portfolio managers investing in both domestic- and foreign-issued bonds use weighted average to obtain portfolio duration. The underlying assumptions of this practice are that foreign and domestic interest rate shifts are perfectly correlated and that exchange rate changes are trivial.

Dym (1991, 1992) demonstrates that Macaulay duration is not a sufficient measure for riskiness of foreign bonds. He proposes an alternative duration measure, which incorporates the local interest rate risk, the exchange rate risk, and their interaction. Chow, Lee, and Solt (1997) show that the correlation (interaction) between the exchange rate and the local interest rate is the predominant factor for changes in US bond returns.

Following Dym (1991, 1992) and Chow et al. (1997), Afik, Jacoby, and Wiener (2018) present the following *ex-ante* two-factor duration model, accounting for both foreign interest rate and exchange rate exposures. Consider a T -year zero-coupon, default-free foreign bond. The value of this bond is subject to three sources of risk related to the foreign exchange rate (x), the domestic riskless rate (r), and the foreign riskless rate (R). For a U.S.-based investor, the foreign exchange rate, x , is determined by the amount of U.S. dollars (domestic currency) paid for a unit of foreign currency. The domestic value of a T -year zero-coupon, default-free foreign bond (P), paying one unit of the foreign currency at maturity, is a function of the parameters x and R . Assuming a flat term structure in both countries, we have:

$$P(x, R) = xe^{-RT}.$$

A simple arithmetic and practical approximation leads to the following expression for the percentage change in the domestic value of the foreign bond:

$$\frac{dP}{P} \cong \frac{dx}{x} - dR \cdot T.$$

The goal is a price elasticity expression with respect to the domestic rate r . Thus, one needs to express dR , a change in foreign interest rate, as a function of dr , the change in the domestic interest rate, and dx , the change in the exchange rate. This can be achieved by a standard (zero-intercept) linear approximation model to describe stochastic processes followed by the dynamics of the foreign short riskless rate:

$$dR = adr + bdx + \varepsilon$$

where a and b are currency-specific constant parameters (that can be estimated empirically), and ε is random noise. Despite its simplicity, this linear representation is supported by its explanatory power when applied to the data. Afik et al. (2018) show that it is highly suitable for a US investor with respect to different currencies.

Combining the above equations results in the following two-factor duration model for foreign bonds:

$$\frac{dP}{P} \cong - \left[T \left(a + b \frac{dx}{dr} \right) - \frac{dx}{dr} \frac{1}{x} \right] dr = -D_{FX} \cdot dr \quad (7)$$

where $D_{FX} \equiv T(a + b \frac{dx}{dr}) - \frac{dx}{dr} \frac{1}{x}$ is the two-factor foreign-exchange risk adjusted duration of the bond.

Noting that $D_x = -\frac{dx}{dr} \frac{1}{x}$ is the elasticity of the foreign-exchange rate with respect to the domestic riskless rate, the two-factor foreign-exchange risk-adjusted duration of the bond, implied by the above equations is given by:

$$D_{FX} = \frac{dR}{dr} T + D_x, \quad (8)$$

where T is the “unadjusted” Macaulay duration of the pure-discount foreign bond.

This model overcomes some prior literature weaknesses such as the model derived by Fooladi, Jacoby, Roberts, and Wiener (2006), where the main difficulty arises from the need to estimate the term dR/dr .

Afik et al. (2018) present a step-by-step model estimation process and apply it to real market data. Since the model ignores sovereign risk, the sample is restricted to high-quality sovereign bonds issued by national governments in five countries: Australia, Canada, Denmark, Great Britain, and Germany. German government bonds serve as a proxy for the Euro countries to avoid the credit-risk premium that affects some of the Euro zone treasuries. The data include daily frequency of 2-, 5-, and 10-year strip rates derived from Bloomberg’s benchmark curves for the period January 2000 to August 2015, and their respective daily exchange rates relative to the U.S. dollar. The analysis shows that the bond price change prediction errors of single-factor models are two- to nine-fold larger than those estimated with the above *two-factor duration model*, depending on bond maturity and currency.

14.9 Stochastic Process Risk: Immunization Complication

Macaulay duration uses yield to maturity as the discount rate, as in Eq. 1. Because yield to maturity discounts all bond cash flows at an identical rate, Macaulay duration implicitly assumes that the interest rates are generated by a stochastic process in which a flat term structure shifts randomly in a parallel fashion so that *all* interest rates change by the same amount. When we assume a different stochastic

process, we obtain a duration measure different from Macaulay duration (Bierwag 1977; Bierwag et al. 1982a). If the actual stochastic process is different from what we assume in obtaining our duration measure, our computed duration will not truly represent the portfolio's risk. In this case, equating the duration measure to the investment horizon will not immunize the portfolio and there will be stochastic process risk. Although immunization is a passive strategy that is not based on an interest rate forecast, it is necessary to predict the stochastic process governing interest rate movements.

A number of researchers have developed strategies for minimizing stochastic process risk and its consequences (for example, Fong and Vasicek 1983, 1984; Bierwag et al. 1987, 1993; Prisman and Shores 1988; Fooladi and Roberts 1992). In particular, Bierwag, Fooladi, and Roberts show that the stochastic process risk is lower when the duration-matched portfolio contains at least one bond that matures at the horizon date. Their argument is reinforced by an empirical study by Soto (2001) who used Spanish government data for the January 1992 to December 1995 period to test the performance of the polynomial duration model. The latter model was derived by Chambers, Carleton, and McEnally (1988) and generalized by Nawalkha and Chambers (1997). Soto (2001) reports that to guarantee the target return (or a return close to that) for a default-free and option-free portfolio, certain constraints of the level, slope, and curvature of the term-structure shift must be satisfied. However, she finds that the exception occurs when a portfolio includes an asset maturing at or near the target horizon date, which is exactly the point of Bierwag et al. (1993). Prior to Soto's (2001) study, Fooladi and Roberts (1992) test the fit of a polynomial term structure for Canadian government bonds and report successful immunization results.

The advantage of including one bond that matures at the target horizon date is also shown in a study by Soto (2004). Using Spanish government bond data for the January 1992 to December 1999 period, she examined the performance of several duration models and a number of risk factors. Soto concludes that the included factors have a more important role in the performance of immunization portfolios than the duration model chosen. She also reports that "only duration-matching bullet and barbell portfolios including a maturity bond perform better."

The importance of forming a portfolio with at least one bond that matures at the horizon date is also confirmed in a study by Nawalkha, Soto, and Zhang (2003) who develop the notion of generalized M-squared and M-vector. Based on these notions, they test the performance of differently modelled immunized portfolios, using U.S. Treasury term structure data for the 1947–1991 period. They conclude that "the pattern of immunization performance across models and horizon lengths collaborates the finding of Fooladi and Roberts (1992) and Bierwag et al. (1993) about the role of the maturity bonds in improving immunization performance." This is also confirmed by the work of Bravo and Pereira da Silva (2006) who apply Portuguese government data. They report that "duration-matching portfolios constrained to include the maturity bond and using a single-factor model provide the best immunization performance overall, particularly in highly volatile term structure environments and shorter holding periods."

In a related criticism of Macaulay duration, Ingersoll, Skelton, and Weil (1978) argue that the assumed stochastic process in the development of single-factor

duration models is inconsistent with equilibrium conditions. The source of arbitrage opportunities is the convexity of the holding period return from immunized portfolios with respect to interest rate shifts. Total value is a convex function of interest rate changes, which for the immunized funds has its minimum at the point of the original rate, r_0 . This means that holding period return is also a convex function of interest rate shifts with the minimum at the original rate. Thus, the larger the shift, the greater the benefit from an interest rate shock. Therefore, in particular in the presence of large shocks to interest rates and or for high coupon bonds, riskless arbitrage would be possible by investors who short zero-coupon bonds for a return of r_0 and invest in other bond portfolios without taking an extra risk.

Although this argument is sound, it does not mar the validity of immunization strategies. Bierwag et al. (1982a) develop an additive stochastic process that is consistent with general equilibrium and for which the holding period return is not a strictly convex function of interest rate shifts. Further, Bierwag (1987b) shows that there is no one-to-one correspondence between a particular duration measure and its underlying stochastic process. Duration measures derived from some disequilibrium processes such as the Fisher–Weil process, the Khang process, and additive and multiplicative processes of Bierwag also correspond to equilibrium processes. This point is also supported by Balbás and Ibáñez (1998) who present general conditions that guarantee a minimum return at the horizon without being inconsistent with equilibrium. Additionally, Bierwag and Roberts (1990) find that examples of equilibrium stochastic processes give rise to duration measures that have been previously derived from disequilibrium stochastic processes such as Fisher–Weil duration.

On the practical side, the riskless-arbitrage argument seemed hypothetical to many practitioners who were aware of the difficulties in taking a large short position in bonds. Practitioners were more concerned that the reality of non-parallel shifts in sloping yield curves could impair the hedges constructed based on Macaulay duration.

The current generation of models incorporates the term structure so that it is no longer the case that duration users must assume a flat term structure. Bierwag, Kaufman, and Toevs (1983), Brennan and Schwartz (1983), and Bierwag et al. (1987, 1993) are a few examples. Current models also allow for non-parallel shifts in the yield curve. These include multi-factor models (Chambers et al. 1988; Nawalkha and Chambers 1997) in which the short and long ends of the yield curve are allowed to shift in opposite directions.⁶

14.10 Effectiveness of Duration-Matched Strategies

Given that duration extensions are numerous, how effective is basic Macaulay duration in the design and implementation of active and passive strategies? Bierwag and Roberts (1990) test the key implication of duration theory for active managers – portfolios with higher durations are predicted to have greater price sensitivity when rates change.

⁶Detailed discussion of these modelling issues is in Bierwag (1977, 1987a).

Constructing portfolios with constant durations using Government of Canada bonds, they measure monthly holding period returns over the period 1963–1986. They find that, as predicted, higher duration portfolios had greater return volatility and that Macaulay duration explained around 80% of the variance in holding period returns.

A number of studies examine the effectiveness of immunization over sample sets of government bonds for the U.S., Canada, and Spain, among others. Fooladi and Roberts (1992) use actual prices for Government of Canada bonds over the period 1963–1986, setting the investment horizon to 5 years and rebalancing every 6 months to maintain duration equal to the time remaining in the investment horizon. Their performance benchmark consists of investing in a bond with maturity matched to the horizon. This maturity-bond strategy involves buying and holding a bond with an initial five-year maturity. Due to stochastic process risk, there will always be cases in which the duration hedging strategy falls short or overshoots the promised return that brings the target proceeds. Their test measures hedging performance so that the better strategy is the one that comes closer most often to the original promised return. They conclude that duration matching allows the formation of effective hedges that outperform non-duration-matched portfolios. These results validate the widespread use of Macaulay duration in measuring risk and in immunization. They also support similar results obtained for U.S. Treasury Securities by Bierwag, Kaufman, Schweitzer, and Toevs (1981) and by Bierwag, Kaufman, and Toevs (1982b).

Beyond establishing the credibility of duration matching as a hedging technique, empirical research has also probed the hedging effectiveness of alternative duration-matching strategies in the face of stochastic process risk. Fong and Vasicek (1983, 1984) propose hedging portfolios designed by constraining M-squared, a measure of cashflow dispersion. Prisman and Shores (1988) and Bierwag, Fooladi, and Roberts (1993) show that the Fong–Vasicek measure does not offer a general solution to minimizing hedging error. The latter paper reinforces results in tests of duration effectiveness discussed earlier which find that stochastic process risk is best controlled by constraining the portfolio to include a maturity bond. This result is replicated for immunization in the Spanish government bond market by Soto (2001) and others (as mentioned in the last section). These include Balbás, Ibáñez, and López (2002) who develop a generalized measure of cash-flow dispersion as immunization risk and report that their empirical tests “support the findings that portfolio minimizing linear measure has the best performance and generally includes maturity bond.”

Going beyond Macaulay duration, Soto (2004), Nawalkha and Soto (2009), among others, examine the increase in hedging effectiveness offered by multi-factor duration models. They establish that, in the absence of the maturity bond constraint, a three-factor model controlling for the level, slope, and term structure curvature works best.

14.11 Use of Financial Futures

The basics of duration analysis can be combined with the use of futures markets instruments for hedging purposes. An investor can hold a long position in a certain security (for example, three-month banker's acceptances) and a short position in a

futures contract written on that security and reduce overall exposure to interest rate risk. This is because, as interest rates change, prices of the security and the futures contract move in the same direction and gains and losses in the long and short positions largely cancel out. The durations of the security held long and of the security underlying the futures contract determine the hedge ratio. When we combine the cash and futures positions, the duration of the overall portfolio may be expressed as:

$$D_P = D_C + D_F(V_F/V_C) \quad (9)$$

where, D_C , D_F , and D_P denote durations of the cash portfolio, futures portfolio, and the overall portfolio, respectively, and V_F and V_C are the values of futures and cash positions, respectively. It should be noted that $V_F = hF$ where F denotes the futures price and h is the number of future contracts per unit of cash portfolio (the hedge ratio).

Bierwag (1987a) shows that, for a perfect hedge, in general the hedge ratio (h) is determined by:

$$h = \frac{1 + r_0^f}{1 + r_0^c} \times \frac{-D_C}{D_F} \times \frac{V(r_0^c)}{F(r_0^f)a} \quad (10)$$

where,

$V(r_0^c)$ = the value of the long position

$F(r_0^f)$ = the future price for one unit of contract

r_0^c = the current yield to maturity of the long asset

r_0^f = the current yield to maturity of the asset underlying the futures contract

a = the derivative r^f with respect to r^c

If the underlying asset is the same and the maturities of asset and future contract are identical, it may be reasonable to assume $a = 1$. Equation (7) shows how an investor can use futures contracts to alter duration of a bond portfolio.

14.12 Duration of Corporate Bonds

Our review of immunization research has so far concentrated on government bonds and ignored default risk. In practice, bond portfolio managers often hold corporate, state, and municipal bonds to enhance yields. This raises the question of how to apply immunization to such portfolios. Simply using Macaulay (or Fisher–Weil) duration for each bond to find portfolio weights will be misleading. Ignoring default risk is equivalent to assuming that we have locked in a higher yield than is possible by immunizing with government bonds alone. The promised return must be adjusted to an expected return reflecting the probability of default.

As Bierwag and Kaufman (1988) argue, in computing duration for non-default-free bonds, in addition to the stochastic process governing interest rate shifts, we must also consider the stochastic process governing the timing of the losses from default. Default alters both a bond's cash flows and their timing. Thus, we cannot immunize a portfolio of non-default-free bonds at its promised rate of return. An interesting question follows. Is it possible to immunize such a portfolio at its (lower) risk-adjusted return using a single-factor duration model?

Fooladi, Roberts and Skinner (1997) answer affirmatively but contend that Macaulay duration is not a true measure of interest rate sensitivity for bonds with default risk. Assuming risk-averse investors, they derive a general expression for duration, which includes terms for default probabilities, expected repayment, and timing of repayment. They illustrate that, under certain circumstances, their general single-factor duration measure is an immunizing measure. They conclude that practical application of duration analysis in immunization calls for employing duration measures that are adjusted for default risk.

Jacoby (2003) extends the model of Fooladi, Roberts, and Skinner (1997) by representing bondholders' preferences with a log-utility function. Accounting for default risk, his duration measure is the sum of the Fisher–Weil duration and the duration of the expected delay between the time of default and actual recovery caused by the default option. Using historical long-term corporate bond default and recovery rates, he numerically simulates his duration measure. His conclusion is that failing to adjust duration for default is costly for high-yield bonds but appears to be trivial for investment-grade bonds.

Since many corporate bonds are callable, Acharya and Carpenter (2002) also use option-pricing technology to derive a valuation framework of callable defaultable bonds. In their model, both interest rates and firm value are stochastic and the call and default decisions are endogenized. With respect to interest-rate sensitivity, as in earlier work by Chance (1990), their model implies that default risk alone reduces the bond's duration. They further show that, everything else being equal, call risk will also shorten bond duration.

The theoretical work of Chance (1990) and Acharya and Carpenter (2002) emphasizes the significance of adjusting Macaulay duration for both default and call risks. Jacoby and Roberts (2003) address the question of the relative importance for duration of these two sources of risk. Using Canadian corporate data, they compare the default- and call-adjusted duration with its Macaulay counterpart. In general, their results support the need for callability adjustment, but fail to uncover any significant impact of default risk for investment-grade callable and defaultable bonds. Their results provide some support for the Acharya and Carpenter (2002) model that predicts an interaction between call and default risks. Jacoby and Roberts demonstrate that during a recessionary period (1991–1994), the call adjustment is less important (but still significant) relative to other periods. This is because bond prices are depressed during recessionary periods, and therefore the incentive to call these bonds arising from lower interest rates is significantly reduced.

Yan et al. (2009) study the effect of default and call risk as well as their interactions on bond duration and basically support the conclusions made by Jacoby

and Roberts. Their findings show that, with few exceptions, default risk shortens duration for risky bonds, call risk always shortens duration of default-free bonds, and the joint effect of call and default risks shortens duration of corporate bonds. Consistent with the work of Jacoby and Roberts (2003), they report that for deep discount bonds, the effect of call risk is milder.

Using factor analysis, Bertocchi, Giacometti, and Zenios (2005) developed a multi-factor model of corporate bond yields and report that the model captures over 98% of the changes in the corporate bond yields. However, as is typical in all factor analysis cases, financial and economic interpretation of these factors are not readily known.

14.13 Macrohedging

This section broadens the application of duration to *macrohedging* addressing interest rate exposure at the macro level of a corporate entity, in particular a financial institution. Macrohedging or balance sheet hedging considers the entire balance sheet and treats both sides as variable. Without a macro approach to hedging, we are forced to take the liability side as given and cannot address asset/liability management.

As with its micro counterpart, macrohedging can be a tool for either passive or active strategies. In a passive (*routine hedging*) strategy, immunization for example, hedging seeks to eliminate interest rate risk completely. In contrast, an active (*selective hedging*) approach leaves some interest rate exposure unhedged seeking to achieve superior returns based on a view of future rates.

14.14 Duration Gap

A duration gap measures the mismatch between assets and liabilities. When the duration gap is zero, the assets and liabilities are perfectly matched so that the financial institution's net worth is immunized against shifts in interest rates. We illustrate duration gap using a simple example involving *one* discount rate for all assets and a financial institution with a single asset, which is financed partly by equity and partly by liabilities. The balance sheet identity requires:

$$A = E + L \quad (11)$$

where A denotes assets, E denotes equity, and L denotes liabilities. Taking the derivative of (11) with respect to the interest rate (single discount rate) and rearranging the terms we obtain:

$$dE/dr = dA/dr - dL/dr \quad (12)$$

For the equity to be unaffected by changes in interest rates, the right-hand side of Eq. (12) must be zero. Multiplying both sides of (12) by a fixed quantity, $(1 + r)/A$, and noting the definition of duration, we obtain

$$\frac{E}{A}D_E = D_A - \frac{L}{A}D_L = D_{Gap} \quad (13)$$

where, D_A , D_L , and D_E denote the durations of assets, liabilities, and equity, respectively. The right-hand side of Eq. (13) is called duration gap, D_{Gap} . A zero duration gap tells us that the equity has zero interest rate exposure.

Restating the equation for duration of equity in terms of Eq. 3, replacing P (for price) with E (for equity), results in Eq. (14).

$$D_E \cong - \left[\frac{\Delta E/E}{\Delta r/(1+r)} \right] \quad (14)$$

Substituting for D_E from (14) into (13) and rearranging the terms results in the following formula for changes in the value of equity as a function of duration gap⁷:

$$\Delta E \cong -D_{Gap} \left[\frac{\Delta r}{(1+r)} \right] \times A \quad (15)$$

This highly useful formula shows how a shift in interest rates impacts the market value of an Financial Institution's equity.⁸ It is set up so that duration gap mathematically plays the same role as duration in the corresponding formula for fixed income securities in Eq. (4). When duration gap is zero, the shares of the FI will not be affected by interest rate shocks; the FI's shares will behave like floating rate bonds with zero duration. The shares of an FI with a positive duration gap will rise when rates fall analogously to a long position in a bond. If an FI has a negative duration gap, its shares will increase in value when rates rise. Holding the shares of such an FI is like taking a short position in a bond.⁹

It follows that, in parallel with fixed income securities, the formula (Eq. 15) has practical implications both for passive management (immunization) and for active management (interest rate speculation). To illustrate, suppose that the management

⁷For simplicity, the derivation assumes away any difference between rates of return for assets and liabilities, r_A and r_L . This may not be strictly true because interest earned on assets is higher than interest paid on liabilities. However, the essence of the argument is not affected by this assumption. In practice, duration gap implementation uses the average of the rates on assets and liabilities.

⁸Following prior research, we choose the change in the market value of equity (E) as the target because maximizing equity value is most likely to be the goal of management and shareholders. Another possible target is E/A (the capital ratio) particularly in the case in which E/A is at the regulatory minimum of 8% and management wishes to immunize against any fall in the ratio. In practice, a financial institution likely targets equity along with other variables. Further discussion of this issue is in Kaufman (1984) and Bierwag and Kaufman (1996).

⁹Some readers may be familiar with the concept of *funding gap*. Funding gap is defined as rate sensitive assets minus rate-sensitive liabilities. It differs from duration gap in two fundamental respects. First, funding gap relates interest rate shifts to the *book value of net income*; duration gap relates rate shifts to the *market value of equity*. Second, funding gap ignores the repricing of long-term assets when rates change. Because of these differences, a positive funding gap corresponds to a negative duration gap.

of an FI regards future interest rate movements as highly uncertain. In this case, the FI should immunize by setting the duration gap to zero. On the other hand, if senior executives predict that rates will fall, the FI should expand its portfolio of longer-term loans financed by short-term deposits increasing D_{Gap} .¹⁰

Central to this strategy is the implied assumption that the difference between convexity of assets and convexity of liabilities (adjusted for capital structure and the ratio of return on assets over return on liabilities) is non-negative. Fooladi and Roberts (2004) show that if this difference, which they call “convexity gap,” is not non-negative, satisfying duration condition is not sufficient for hedging against interest rate risk.

To reduce adjustment time, and save on transactions costs, FIs adjust duration gaps using off-balance-sheet positions in derivative securities such as interest rate futures, interest rate options, and swaps. Bierwag (1977) shows that to find the proper hedge ratio for futures hedging, we simply substitute D_{Gap} for D_c in Eq. (8) (that is used for constructing hedged positions in bond portfolios).

Constructing hedged positions is complicated by pre-payment options built into mortgage loans and mortgage-backed securities. If rates rise from historic lows, pre-payment options will become deep out of the money causing asset durations to lengthen substantially (Nawalkha and Soto 2009).

14.15 Other Applications of Duration Gaps

Duration gap also has applications to managing the balance sheets of life insurance companies and pension funds, and even other entities such as nonfinancial corporations and governments.

Life insurance companies were the first class of financial institutions to implement duration matching.¹¹ Policy reserves are the main liability of a life-insurance company and represent the expected present value of future liabilities under life policies. The typical life insurance company invests the bulk of its assets in bonds and mortgages. This leads to a constitutionally negative duration gap as future policy liabilities generally have a longer duration than bonds and mortgages. To address the resulting interest rate exposure, during the 1980s life-insurance companies increased their investments in mortgages and other positions in real estate. The recession and real-estate collapse of the early 1990s led to the insolvency of several life insurance companies. Today, well-managed life insurance companies recognize that off-balance-sheet positions in futures and other derivatives offer an attractive way to hedge an imbalanced duration gap without the risks that come with large positions in real estate.

¹⁰More generally, academic research supports the view that FI shares move in the direction expected by the theory of duration gaps. For an example, see Flannery and James (1984).

¹¹Early researchers on duration (Macaulay 1938; Redington 1952, for example) were actuaries who published their results in actuarial journals.

Pension plans come in two types: defined benefit and defined contribution. The balance sheet of a defined benefit pension plan, like that of a life insurance company, has a constitutional imbalance in duration. Given that an average employee may retire 20 years from today, and then live for another 20 years, the duration of the pension liability is generally longer than the duration of the asset portfolio invested in equities and bonds. As a result, in the 1990s, many defined benefit pension plans were increasing their equity exposures and taking equity positions in real estate. Bodie (1996) shows that this leaves pension funds exposed to mismatched exposures to interest rate and market risks.

Beginning in 2000, sharp declines in both stock prices and long-term interest rates created a “Perfect Storm” for defined benefit pension funds (Zion and Carache 2002). As a result, most plans have switched to the defined contribution format.¹² Many pension plans offer some form of indexation of retirement benefits to compensate for inflation that occurs after employees retire. To address inflation risk, pension funds can immunize all or part of their liabilities against interest rate risk using macro or micro hedging, and then add an inflation hedge. The portfolio shifts to equities and real estate investments discussed earlier offer a potential inflation hedge. Another attractive possibility is Treasury Inflation-Protected Securities (TIPS). These bonds offer indexation of principal and interest, and thus have approximately zero inflation duration. Note, however, that TIPS provide only partial indexation. While coupons are fully indexed, the principal of TIPS is only indexed for positive inflation (see Jacoby and Shiller 2008).¹³

Duration gaps can also be used to hedge the equity of entities such as nonfinancial corporations and governments against interest rate fluctuations. For example, in an innovative application of duration analysis, at the end of the previous millennium the New Zealand government explored how this tool could help guide restructuring of its liabilities (see Bierwag, Fooladi, and Roberts (2000), who were the first to make the distinction between real and inflation duration).

Fooladi, Jacoby, and Jin (2021) demonstrate a strategy to hedge the equity of an entity in the presence of inflation. They note that duration depends on the source of the change in the interest rate, which could result from shifts in both the real and expected inflation rate. They further show that the nominal duration of an instrument is the sum of its real interest rate elasticity (real duration) and its elasticity with respect to expected inflation (inflation duration). To hedge the equity of an entity against interest-rate volatility, Fooladi et al. (2021) propose a joint hedge strategy that simultaneously protects against shifts in both real interest rates and expected inflation rates. This strategy involves jointly setting the real-duration gap and the

¹²According to the U.S. Bureau of Labor Statistics, in 2019 only 16% of private-sector employees were enrolled in a defined benefit plan (see: www.bls.gov/ncs/ebs/factsheet/defined-benefit-frozen-plans.pdf)

¹³In case of deflation, the principal amount cannot be indexed to a level lower than its unadjusted principal.

inflation-duration gap to zero. This requires that entities simultaneously adjust their debt cash-flow structure and its maturity.¹⁴

Acknowledgments The authors gratefully acknowledge the support of the Social Sciences and Humanities Research Council of Canada. In addition, Iraj Fooladi received support from the Douglas C. Mackay Fund at Dalhousie.

References

- Acharya, V.V., and J.N. Carpenter. 2002. Corporate bond valuation and hedging with stochastic interest rates and endogenous bankruptcy. *Review of Financial Studies* 15: 1355–1383.
- Afik, Z., G. Jacoby, and Z. Wiener. 2018. Duration and globalization. *The Journal of Fixed Income* 28 (2): 31–43.
- Balbás, A., and A. Ibáñez. 1998. When can you immunize a bond portfolio? *Journal of Banking & Finance* 22: 1571–1595.
- Balbás, A., A. Ibáñez, and S. López. 2002. Dispersion measures as immunization risk measures. *Journal of Banking & Finance* 26: 1229–1244.
- Bertocchi, M., R. Giacometti, and S.A. Zenios. 2005. Risk factor analysis and portfolio immunization in the corporate bond market. *European Journal of Operational Research* 161: 348–363.
- Bierwag, G.O. 1977. Immunization, duration, and the term structure of interest rates. *Journal of Financial and Quantitative Analysis* 12 (5): 725–741.
- . 1987a. *Duration analysis, managing interest rate risk*. Cambridge, MA: Ballinger Publishing.
- . 1987b. Bond returns, discrete stochastic processes, and duration. *Journal of Financial Research* 10 (3): 191–209.
- Bierwag, G.O., and I. Fooladi. 2006. Duration analysis: An historical perspective. *Journal of Applied Finance* 16 (2): 144–160.
- Bierwag, G.O., and G.G. Kaufman. 1977. Coping with the risk of interest-rate fluctuations: A note. *Journal of Business* 50 (3): 364–370.
- . 1988. Duration of non-default free securities. *Financial Analysts Journal* 44 (4): 39–46.
- . 1996. Managing interest rate risk with duration gaps to achieve multiple targets. *Journal of Financial Engineering* 5: 53–73.
- Bierwag, G.O., and C. Khang. 1979. An immunization strategy is a minimax strategy. *Journal of Finance* 34 (2): 389–399.
- Bierwag, G.O., and G.S. Roberts. 1990. Single factor duration models: Canadian tests. *Journal of Financial Research* 13: 23–38.
- Bierwag, G.O., G.G. Kaufman, R. Schweitzer, and A. Toevs. 1981. The art of risk management in bond portfolios. *Journal of Portfolio Management* 7: 27–36.
- Bierwag, G.O., G.G. Kaufman, and A. Toevs. 1982a. Single factor duration models in a discrete general equilibrium framework. *Journal of Finance* 37: 325–338.
- . (1982b). *Empirical tests of alternative single factor duration models*. Paper presented at Western Finance Association, Portland, Oregon (June).
- . 1983. Immunization strategies for funding multiple liabilities. *Journal of Financial and Quantitative Analysis* 18: 113–124.

¹⁴Note that the inflation-indexation schemes will affect an instrument's value and elasticity. Fooladi et al. (2021) demonstrate that, when imperfect indexation implies under (over) protection against inflation, the instrument will have a positive (negative) expected-inflation duration, and its value will be negatively (positively) affected by inflation.

- Bierwag, G.O., G.G. Kaufman, and C. Latta. 1987. Bond portfolio immunization: Tests of maturity, one- and two-factor duration matching strategies. *Financial Review* 22 (2): 203–219.
- Bierwag, G.O., I. Fooladi, and G. Roberts. 1993a. Designing and immunization portfolio: Is M -squared the key? *Journal of Banking & Finance* 17 (6): 1147–1170.
- . 1993b. Designing and immunization portfolio: Is M -squared the key? *Journal of Banking & Finance* 17: 1147–1170.
- . 2000. Risk management with duration: Potential and limitations. *Canadian Journal of Administrative Sciences* 17 (2): 126–142.
- Bodie, Z. 1996. What the pension benefit guaranty corporation can learn from the Federal Savings and Loan Insurance Corporation. *Journal of Financial Services Research* 10: 83–100.
- Bravo, J.M., and C.M.P. Da Silva. 2006. Immunization using a stochastic-process independent multi-factor model: The Portuguese experience. *Journal of Banking & Finance* 30: 133–156.
- Brennan, M.J., and E.S. Schwartz. 1983. Duration, bond pricing, and portfolio management, in innovations in bond portfolio management, duration analysis and immunization. In *Proceedings from the Ashland conference, 1981*, ed. G.G. Kaufman, G.O. Bierwag, and A. Toevs. Greenwich: JAI Press.
- Chambers, D.R., W.T. Carleton, and R.W. McEnally. 1988. Immunizing default-free bond portfolios with a duration vector. *Journal of Financial and Quantitative Analysis* 23 (1): 89–104.
- Chance, D.M. 1990. Default risk and the duration of zero coupon bonds. *Journal of Finance* 45 (1): 265–274.
- Chow, E.H., W.Y. Lee, and M.E. Solt. 1997. The exchange-rate risk exposure of asset returns. *Journal of Business* 70 (1): 105–123.
- Cornett, M.M., and A. Saunders. 2008. *Financial institutions management*. 6th ed. Boston: Irwin McGraw Hill, Chapters 8, 9 and 18.
- Diaz, A., M.O. González, El Navarro, and F.S. Skinner. 2009. An evaluation of contingent immunization. *Journal of Banking & Finance* 33: 1874–1883.
- Dym, S.I. 1991. Measuring the risk of foreign bonds. *Journal of Portfolio Management* 17 (2): 56–61.
- Dym, S. 1992. Global and local components of foreign bond risk. *Financial Analysts Journal* 48 (2): 83–91.
- Fisher, L., and R.L. Weil. 1971. Coping with the risk of market-rate fluctuations: Returns to bondholders from naïve and optimal strategies. *Journal of Business* 44 (4): 408–431.
- Flannery, M.J., and C.M. James. 1984. The effect of interest rate changes on the common stock returns of financial institutions. *Journal of Finance* 39: 1141–1153.
- Fong, H.G., and O. Vasicek. 1983. Return maximization for immunized portfolios, in innovations in bond portfolio management, duration analysis and immunization. In *The Ashland conference, 1981*, ed. G.G. Kaufman, G.O. Bierwag, and A. Toevs. Greenwich: JAI Press.
- Fong, H.G., and O.A. Vasicek. 1984. A risk minimizing strategy for portfolio immunization. *Journal of Finance* 39 (5): 1541–1546.
- Fooladi, I., and G. Roberts. 1992. Portfolio immunization: Canadian tests. *Journal of Economics and Business* 44 (1): 3–17.
- . 2004. Macrohedging for financial institutions: Beyond duration. *Journal of Applied Finance* 14 (1): 11–19.
- Fooladi, I., G. Roberts, and F. Skinner. 1997. Duration for bonds with default risk. *Journal of Banking and Finance* 21 (1): 1–16.
- Fooladi, I.J., G. Jacoby, G.S. Roberts, and Z. Wiener. 2006. Domestic elasticity of default-free foreign bonds. *Journal of Applied Finance* 16 (2): 174–182.
- Fooladi, I.J., G. Jacoby, and L. Jin. 2021. Real duration and inflation duration: A cross country perspective on a multidimensional hedging strategy. *Journal of International Financial Markets, Institutions and Money* 70: 101265.
- Hicks, J.R. 1939. *Value and capital*. 2nd ed. Oxford: Clarendon Press. 1946.
- Hopewell, M.H., and G.G. Kaufman. 1973. Bond price volatility and term to maturity: A generalized respecification. *American Economic Review* 63 (4): 749–753.

- Ingersoll, J.E., J. Skelton, and R.L. Weil. 1978. Duration forty years later. *Journal of Financial and Quantitative Analysis* 13 (4): 621–650.
- Jacoby, G. 2003. A duration model for defaultable bonds. *Journal of Financial Research* 26: 129–146.
- Jacoby, G., and G. Roberts. 2003. Default- and call-adjusted duration for corporate bonds. *Journal of Banking & Finance* 27: 2297–2321.
- Jacoby, G., and I. Shiller. 2008. Duration and pricing of TIPS. *Journal of Fixed Income* 18 (2): 71–84.
- Kaufman, G.G. 1984. Measuring and managing interest rate risk: A primer. *Economic Perspectives* 8: 16–29.
- Leibowitz, M.L., and A. Weinberger. 1981. The uses of contingent immunization. *Journal of Portfolio Management* 8 (1): 51–55.
- . 1982. Contingent immunization, part I: Risk control procedures. *Financial Analysts Journal* 38 (6): 17–32.
- . 1983. Contingent immunization, part II: Problem areas. *Financial Analysts Journal* 39 (1): 35–50.
- Macaulay, F. R. 1938. *Some theoretical problems suggested by the movement of interest rates, bonds, yields, and stock prices in the United States since 1856*. National Bureau of Economic Research. <https://www.nber.org/books-and-chapters/some-theoretical-problems-suggested-movements-interest-rates-bond-yields-and-stock-prices-united>
- Morgan, J.P. 1994. *Introduction to riskmetrics*. New York: J.P. Morgan.
- Nawalkha, S.K., and G.M. Soto. 2009. Term structure estimation. Available at: <https://ssrn.com/abstract=1096182>
- Nawalkha, S.K., and D.R. Chambers. 1997. The *M*-vector model: Derivation and testing of extensions to *M*-square. *Journal of Portfolio Management* 23 (2): 92–98.
- Nawalkha, S.K., G.M. Soto, and J. Zhang. 2003. Generalized *M*-vector models for hedging interest rate risk. *Journal of Banking & Finance* 27: 1581–1604.
- Oliveira, L., J.P.V. Nunes, and L. Malcata. 2014. The performance of deterministic and stochastic interest rate risk measures: Another question of dimensions? *Portuguese Economic Journal* 13: 141–165.
- Ortobelli, S., S. Vitali, M. Cassader, and T. Tichy. 2018. Portfolio selection strategy for fixed income markets with immunization on average. *Annals of Operations Research* 260: 395–415.
- Prisman, E. 1986. Immunization as a maxmin strategy: A new look. *Journal of Banking & Finance* 10 (4): 491–509.
- Prisman, E.Z., and M.R. Shores. 1988. Duration measures for specific term structure estimations and applications to bond portfolio immunization. *Journal of Banking & Finance* 12 (3): 493–504.
- Redington, F.M. 1952. Review of the principle of life-office valuations. *Journal of the Institute of Actuaries* 78: 286–340.
- Samuelson, P.A. 1945. The effects of interest rate increases on the banking system. *American Economic Review* 35 (1): 16–27.
- Soto, G.G. 2001. Immunization derived from a polynomial duration vector in the Spanish bond market. *Journal of Banking & Finance* 25: 1037–1057.
- Soto, G.M. 2004. Duration models and IRR management: A question of dimensions? *Journal of Banking & Finance* 28: 1089–1110.
- Yan, A.X., S. Liu, C. Wu, and B. Anderson. 2009. The effects of default and call risk on bond duration. *Journal of Banking & Finance* 33 (9): 1700–1708.
- Zhu, W., C.H. Zhang, Q. Liu, and S.S. Zhu. 2018. Incorporating convexity in bond portfolio immunization using multifactor model: A semidefinite programming approach. *Journal of the Operations Research Society of China* 6: 3–23.
- Zion, D., and B. Carache. 2002. *The magic of pension accounting*. Boston: Credit Suisse First.



Aron Gottesman

Abstract

Loan contract terms refer to the price and nonprice terms associated with a corporate loan deal between a borrower and a lender or a syndicate of lenders. The specification of loan contract terms differs across loans. These differences are attributable to the tradeoffs between values of loan contract terms that the borrower chooses when negotiating the loan contract, as well as the purpose of the loan and borrower and lending syndicate characteristics. Methodological issues that arise when investigating the relations between loan contract terms include allowing for loan contract terms that are determined simultaneously and accurately estimating credit risk.

Keywords

Collateral · Corporate loans · Credit risk · Credit spread · Loan contract terms · Loan pricing corporation · Maturity · Simultaneity · Syndicate size · Syndicated loans

15.1 Introduction

Loan Contract Terms refer to the price and non-price terms associated with a corporate loan deal between a borrower and a lender or a syndicate of lenders. Corporate loan deals are composed of one or more loans, designated *loan facilities*, and loan contract terms can vary across facilities. Price terms include the rate spread over the risk-free rate, typically the prime rate or LIBOR, and fees such as upfront, annual, cancellation, and commitment fees. Nonprice terms include deal and facility size, maturity, collateral, financial and nonfinancial covenants, and performance

A. Gottesman (✉)

Department of Finance and Graduate Economics, Lubin School of Business, Pace University,
New York, NY, USA

e-mail: agottesman@pace.edu

pricing covenants. Other characteristics across which loans differ include whether the loan is a revolver or term loan, the seniority of the loan, and the size and concentration of the lending syndicate, among others.

The specification of loan contract terms differs across loans. These differences are attributable to the tradeoffs between values of loan contract terms that the borrower chooses when negotiating the loan contract, as well as the purpose of the loan and borrower and lending syndicate characteristics. Melnik and Plaut (1986) model loan commitment contracts as a “package” of negotiated terms. In their model, the loan commitment contract is described by the vector $B[L^*, T, m, k, C]$, where L^* is the amount of credit the lender is willing to provide, T is the maturity of the contract, m is the rate spread, k is the loan commitment fee rate, and C is the collateral. Borrowers can choose to tradeoff less favorable specification of some contract terms in exchange for more favorable specification of other contract terms. Melnik and Plaut (1986) empirically test for the existence of such tradeoffs through investigating whether loan commitment size is related to other loan contract features, and find support for the hypotheses that lenders are willing to provide a larger loan commitment in exchange for a higher spread or more collateral. They also find support for the hypothesis that lenders are willing to provide larger loan commitments in exchange for a continuing customer relationship. As well, Melnik and Plaut (1986) identify a positive relation between loan commitment size and firm characteristics such as proxies for firm credit rating and firm size.

While Melnik and Plaut’s (1986) study provides important early insight into the relation between loan contract terms, subsequent studies contribute to a more complete understanding. One relation that has received attention is the relation between loan spreads and maturity. Two competing hypotheses explain the nature of the relation. The *tradeoff hypothesis* forecasts a positive relation between corporate loan spreads and maturity, while the *credit quality hypothesis* forecasts a negative relation. The positive relation forecasted by the tradeoff hypothesis is based on the observation that, *ceteris parabis*, borrowers prefer to borrow for longer periods to avoid the costs associated with liquidation at maturity, while lenders prefer to lend for shorter periods to avoid agency problems. The negative relation forecasted by the credit quality hypothesis is based on the argument that lenders direct riskier borrowers to shorter-maturity loans, and less-risky borrowers to longer-maturity loans. Because less-risky borrowers are less likely to default, the corporate loan spreads they pay are lower than the spreads paid by riskier borrowers, hence the relation between loan spreads and maturity is forecasted to be negative.

Some empirical evidence identifies a negative relation between loan spreads and maturity, which supports the credit quality hypothesis. Strahan (1999) performs regression estimation of measures of spread against maturity and other regressors, and identifies a statistically significant negative coefficient associated with his measure of maturity for both lines of credit and term loans. Dennis et al. (2000) identify a negative relation as well. But there is also evidence that longer maturity loans are associated with higher spreads (Helwege and Turner 1999; Coleman et al. 2002), supportive of the tradeoff hypothesis. Gottesman and Roberts (2004) argue that both hypotheses can coexist: the credit quality hypothesis at the portfolio level, and the tradeoff hypothesis at the individual firm level. Gottesman and Roberts

(2004) test a matched pair sample consisting of longer and shorter maturity loan facilities between identical lender syndicates and individual borrowers. Both loan facility elements of each matched pair are segments of identical larger loan deals; hence firm and temporal characteristics are controlled. Through the use of these controls any effects associated with the credit quality hypothesis are eliminated, as both elements of each matched pair are associated with the same firm, and, therefore, are characterized by identical credit quality. Gottesman and Roberts (2004) identify a positive relation between loan spreads and maturity using their methodology, and argue that the tradeoff hypothesis is supported at the firm level, while the credit quality hypothesis describes reality at the loan portfolio level.

Another relation that has received attention is the relation between loan spreads and collateralization. There is extensive evidence that loans that are collateralized are associated with higher spreads than noncollateralized loans (Berger and Udell 1990, 1995; Dennis et al. 2000; John et al. 2003; Gottesman and Roberts 2005). Superficially, these finding are odd: shouldn't collateralization reduce the risk associated with the loan, and therefore lead to lower spreads? One explanation for the existence of higher spreads for collateralized loans is that riskier borrowers are more likely to be forced by lenders to collateralize than less risky borrowers, as suggested in theoretical models and empirical papers. Hence, the higher spreads associated with collateralized loans arise because of the riskier nature of these borrowers, even after the risk-reducing effects of collateralization (Berger and Udell, 1990; Pozzolo, 2002). An alternative explanation for the higher spreads associated with collateralized loans is unrelated to the risk characteristics of the borrower; instead, John et al.'s (2003) management-consumption hypothesis argues that the higher spreads are the result of agency problem associated with collateralized debt. Support for this hypothesis comes from empirical evidence of higher *ex ante* spreads (e.g. John et al. 2003; Gottesman and Roberts 2005).

Empirical tests identify a number of additional relations between loan contract terms. For example, there is evidence that larger and less leveraged firms are more likely to borrow revolving loans rather than term loan (Coleman et al. 2002). Further, loans that include a performance-pricing covenant have significant lower spreads than loans without such a covenant (Asquith et al. 2002), though there is also evidence that this result is limited to accounting-based performance pricing covenants (Panyagometh et al. 2004). Accounting-based performance-pricing covenants are associated with collateralization, longer maturity, and riskier loans (Doyle 2003). There is evidence of a complementary pattern between performance pricing and other covenant provisions, though performance-pricing covenants are designed to deal with the scenario where the borrower's credit improves, while other covenants are designed for scenarios where credit deteriorates (Beatty et al. 2002).

15.2 Characteristics of the Lending Syndicate

Corporate loans are provided by either a sole lender or by a syndicate of lenders; indeed, syndicates of lenders provide a large proportion of corporate loans. The characteristics of the lending syndicate are important to both the lender and the borrower, and a lending

syndicate structure that is optimal for the lenders maybe suboptimal for the borrower. We therefore expect tradeoffs between the syndicate structure and other loan contract terms. The arranging bank in a syndicate plays an important role in influencing syndicate size, concentration, and negotiated loan contract terms. As Lee and Mullineaux (2001) discuss, arranging banks control the size and concentration of the syndicate in a number of ways. First, the arranger chooses which lenders to invite into the syndicate. Second, the arranger specifies participation bracket size and fee. Third, the arranger can close the syndication before the end of the offering period.

Lee and Mullineaux (2001) provide arguments as to why syndicate size is important. Larger syndicates can be costly, as unanimous agreement by all participants is required to permit change to the original loan agreement. Hence, should the borrower face financial distress, larger syndicates require costlier renegotiations and are more likely to result in failure to reach unanimous agreement. Because riskier firms are more likely to face financial distress, smaller syndicates are highly desirable for loans to riskier borrowers. Yet the arranging bank may prefer larger syndicates as it allows them to provide participation opportunities to other lenders.

Borrowers may prefer syndicated loans to avoid situations where a sole-lender monopolizes proprietary information about the borrower. As Boot (2000) notes, banks can use their monopoly over proprietary information about the borrower to charge a higher rate than would be expected in a competitive environment (the hold-up problem). One solution for the borrower is to engage in multiple bank relationships and to ensure the availability of competing sources of loans (von Thadden 1992). Syndicated loans can be perceived as a source of multiple relationships. Note, however, that the more the lenders, the more likely that the proprietary information will be leaked. Therefore, Bhattacharya and Chiesa (1995) contend that a firm will form less relationships if it holds valuable proprietary information that it does not wish to leak. Hence, there are tradeoffs associated with multiple banking relationships as well. Empirical evidence suggests that a relationship with a single lender is associated with superior credit availability. There is also mixed empirical evidence regarding the interaction between loan rates and the number of bank relationships in which the borrower is engaged. As for concentration, concentrated loan share gives participants the incentives to monitor and renegotiate in good faith, and is less likely to result in free riding. This suggests that concentrated syndicates are particularly desirable when there are information asymmetries and potential agency issues that require monitoring. Yet participants may wish to limit their exposure, particularly for loans to risky borrowers.

Lee and Mullineaux (2001) perform empirical tests and find that syndicate size is positively related to the information available about the borrower, the term to maturity, and the arranging bank's reputation. Syndicate concentration is positively related to information asymmetry and to the presence of security. Concentration is negatively related to borrower credit quality and lead bank reputation. As well, syndicate size is larger when resale activities are limited and less concentrated. Dennis and Mullineaux (2000) and Jones et al. (2000) find that the share of the

syndicated loan held by the arranging bank is negatively related to loan quality. Esty and Megginson (2003) find that syndicate size is larger and more diffused in countries where lenders cannot rely on legal enforcement mechanisms.

15.3 Methodological Issues

15.3.1 Simultaneity

Dennis et al. (2000) criticize the empirical literature on loan contract terms, arguing that studies that focus on single contract features ignore the econometric issues that arise if contract features are determined simultaneously. For example, Dennis et al. (2000) note that maturity and collateral may be related to common exogenous factors such as credit quality or agency costs. While the simultaneity issue can be resolved through excluding other loan contract terms from OLS estimation, such an approach does not permit analysis of the tradeoff across loan contract terms. To account for simultaneity, Dennis et al. (2000) perform their tests through estimating a simultaneous equation model using two-stage least squares (2SLS) estimation, for a sample of revolver loans, specified as follows:

$$\text{Duration} = \gamma_{DS}\text{Secured} + \beta_1 X_1 + e_1 \quad (1)$$

$$\text{Secured} = \gamma_{SD}\text{Duration} + \beta_2 X_2 + e_2 \quad (2)$$

$$\text{All-In-Spread} = \gamma_{AD}\text{Duration} + \gamma_{AS}\text{Secured} + \gamma_{AC}\text{Comfee} + \beta_3 X_3 + e_3 \quad (3)$$

$$\text{Comfee} = \gamma_{CD}\text{Duration} + \gamma_{CS}\text{Secured} + \gamma_{CA}\text{All-In-Spread} + \beta_4 X_4 + e_4 \quad (4)$$

where *duration* is maturity; *secured* is a collateralization dummy; *all-in-spread* is the basis point coupon spread over LIBOR plus the annual fee and upfront fee, spread over the life of the loan; *comfee* is the commitment fee; and **X** is a vector of other control variables that measure firm characteristics such as risk and size, and loan characteristics such as loan purpose and structure. This model captures the tradeoffs suggested by Melnik and Plaut (1986) through two bi-directional relations: between duration and security, and between spreads and fees. It also allows the values of duration and security to influence spreads and fees.

The 2SLS estimation performed by Dennis et al. (2000) provides evidence of a positive relation between maturity and collateralization, and between all-in-spreads and commitment fees. As noted earlier, they also find evidence that spreads are negatively related to maturity and positively related to collateralization. To demonstrate that accounting for simultaneity is critical, Dennis et al. (2000) repeat their estimation using single equation estimation and fail to find evidence of the relation between maturity and collateralization. Further, single equation estimation results in evidence of a positive relation between commitment fee and both maturity and

collateralization. These results differ from the results when 2SLS estimation is used. Dennis et al. (2000) use the differences between the results for single equation and 2SLS estimation as evidence that ignoring simultaneity can "...produce potentially biased and inconsistent estimates of the relationships." (p. 107).

15.4 Measures of Risk

The credit riskiness of the borrower strongly influences the negotiated package of terms. Most obviously, we expect a lender to demand a higher spread from a borrower with a higher probability of default, to compensate for the additional risk with which the lender is burdened. Riskiness also influences important loan contract terms such as maturity and collateralization, as discussed earlier. The influence of credit riskiness on loan contract terms requires that it be controlled when relating loan contract terms to each other; yet riskiness of the borrower is often difficult to estimate.

One measure that is frequently used to control for the borrower's riskiness is the borrower's credit rating. However, credit ratings are an inadequate control for risk, as they do not provide useful information about short- and medium-term likelihood of default. Ratings tend to overestimate risk when the economy is strong and underestimate risk when the economy is weak, due to systematic variations in the relation between ratings and risk. This effect is further exacerbated by change in the risk-free rate of interest as the economy changes. Other measures of credit risk based on long-term averages, such as the variance of earnings, are also inadequate for similar reasons.

One alternative is to use an options theoretic approach to estimate default risk in the spirit of Merton (1974). A relatively easy-to-implement method of estimating the implied probability of default is described by Saunders and Allen (2002, Chapter 4) and Allen and Peristiani (2004). The implied default probability is $N(-DD_{it})$, where

$$DD_{it} = \frac{\ln\left(\frac{V_{Ait}}{L_{it}}\right) + T(r_t + 0.5\sigma_{Ait}^2)}{\sigma_{Ait}\sqrt{T}} \quad (5)$$

In this equation, borrower i 's asset value and asset volatility at time t , V_{Ait} and σ_{Ait} , are identified through solving the following system of nonlinear equations:

$$V_{Eit} = V_{Ait}N(DD_{it}) - e^{-r_t T}L_{it}N(DD_{it} - \sigma_{Ait}\sqrt{T}) \quad (6)$$

$$\sigma_{Eit} = \frac{V_{Ait}}{V_{Eit}}N(DD_{it})\sigma_{Ait}, \quad (7)$$

where V_{Eit} is the market value of borrower i 's equity at time t , L_{it} the borrower's debt, r_t the risk-free rate, σ_{Eit} the borrower i 's equity volatility at time t , T the period,

and $N()$ the normal distribution. Allen and Peristiani note that the implied default probability estimated using the above methodology does not exactly correspond to the actual probability due to the normality assumption. However, they argue that this measure reflects variations in the probability of default. These characteristics make it a useful companion to measures of average long-term risk such as credit ratings.

Notes

1. We primarily focus on the literature related to private loans in this discussion.
2. Many studies use data from the Loan Pricing Corporation's (LPC) Dealscan database, which details price and nonprice loan contract terms associated with syndicated loans. The LPC Dealscan database reports a number of measures of spread for each loan facility including the prime spread; the LIBOR spread; and measures that combine spread and fees.
3. The values of loan contract terms can also be influenced by macroeconomic factors, as well as loan market factors such as regulation and competitiveness. For example, see Berger and Hannan (1989), Petersen and Rajan (1995), Hannan (1997), Covitz and Heitfield (1999), Boot and Thakor (2000), Beck et al. (2004), among many others.
4. Other early papers that relate demand for credit to loan contract terms include Azzi and Cox (1976), Arzac et al. (1981), and Koskela (1983).
5. Agency problems associated with longer maturity loans include asset substitution and underinvestment. See Myers (1977) and Barnea et al. (1980). Also see signaling arguments in Flannery (1986) and Kale and Noe (1990), which suggest that less risky borrowers will choose shorter loans.
6. Dennis et al. (2000) review and develop a number of hypotheses that relate maturity and collateral to other borrower characteristics besides credit quality; for example, their tax hypothesis predicts that maturity is inversely related to the firm's marginal tax rate, and positively related to the slope of the yield curve.
7. Also see Berger and Udell (1990) and Guedes and Opler (1996).
8. See Boot et al. (1991), Bester (1994), Coco (1999), Hester (1979), Berger and Udell (1990, 1995), Carey et al. (1998), and Harhoff and Korting (1998). Note that others argue that less risky borrowers are more likely to collateralize. See Bester (1985, 1987), Chan and Kanatas (1985) and Besanko and Thakor (1987).
9. Also see Rajan (1992).
10. Degryse and Ongena (2001) empirically investigated publicly listed Norwegian firms, and found that firm profitability is negatively related to the number of relationships that the firm has with banks. They interpret this to suggest that young firms begin with bilateral relationships, and remain with the bank if successful. Mediocre firms, on the other hand, develop multiple banking relationships.
11. Also see Bolton and Scharfstein (1996) and Detragiache et al. (2000).
12. See Degryse and Ongena (2001).
13. See Standard and Poor's (2004) and Treacy and Carey (1998).

References

- Allen, L. and Peristiani, S. (2004). Conflicts of interest in merger advisory services. Working Paper, Baruch College, May.
- Arzac, E.R., R.A. Schwartz, and D.K. Whitcomb. 1981. A theory and test of credit rationing: some further results. *American Economic Review* 71: 735–777.
- Asquith, P., Beatty, A., and Weber, J. (2002). \Performance pricing in debt contracts. Working Paper, Penn State University and MIT.
- Azzi, C.F., and J.C. Cox. 1976. A theory and test of credit rationing: comment. *American Economic Review* 66: 911–917.
- Barnea, A., R. Haugen, and L. Senbet. 1980. A rationale for debt maturity structure and call provisions in an agency theoretic framework. *Journal of Finance* 35: 1223–1234.
- Beatty, A., Dichev, I.D., and Weber, J. (2002). The role and characteristics of accounting-based performance pricing in private debt contracts. Working Paper, Penn State University, University of Michigan, and MIT.
- Beck, T., A. Demircuc-Kunt, and V. Maksimovic. 2004. Bank competition and access to finance: international evidence. *Journal of Money, Credit, and Banking* 36: 627–649.
- Berger, A.N., and T.H. Hannan. 1989. The price–concentration relationship in banking. *The Review of Economics and Statistics* 71: 291–299.
- Berger, A.N., and G.F. Udell. 1990. Collateral, loan quality and bank risk. *Journal of Monetary Economics* 25: 21–42.
- . 1995. Relationship lending and lines of credit in small firm finance. *Journal of Business* 68: 351–381.
- Besanko, D., and A.V. Thakor. 1987. Collateral and rationing: sorting equilibria in monopolistic and competitive credit markets. *International Economic Review* 28: 671–690.
- Bester, H. 1985. Screening vs. rationing in credit markets with imperfect information. *American Economic Review* 57: 850–855.
- . 1987. The role of collateral in credit markets with imperfect information. *European Economic Review* 31: 887–899.
- . 1994. The role of collateral in a model of debt renegotiation. *Journal of Money, Credit, and Banking* 26: 72–86.
- Bhattacharya, S., and G. Chiesa. 1995. Proprietary information, financial intermediation, and research incentives. *Journal of Financial Intermediation* 4: 328–357.
- Bolton, P., and D. Scharfstein. 1996. Optimal debt structure and the number of creditors. *Journal of Political Economy* 104: 1–25.
- Boot, A.W. 2000. Relationship banking: what do we know? *Journal of Financial Intermediation* 9: 7–25.
- Boot, A.W., and A.V. Thakor. 2000. Can relationship banking survive competition? *Journal of Finance* 55: 679–713.
- Boot, A.W., A.V. Thakor, and G.F. Udell. 1991. Collateralized lending and default risk: equilibrium analysis, policy implications and empirical results. *Economic Journal* 101: 458–472.
- Carey, M., M. Post, and S.A. Sharpe. 1998. Does corporate lending by banks and finance companies differ? Evidence on specialization in private debt contracting. *Journal of Finance* 53: 845–878.
- Chan, Y., and G. Kanatas. 1985. Asymmetric valuations and the role of collateral in loan agreements. *Journal of Money, Credit, and Banking* 17: 84–95.
- Coco, G. 1999. Collateral and heterogeneity in risk attitudes and credit market equilibrium. *European Economic Review* 43: 559–574.
- Coleman, A.D.F., Esho, N., and Sharpe, I.G. (2002). Do bank characteristics influence loan contract terms. Australian Prudential Regulation Authority Working Paper.

- Covitz, D. and Heitfield, E. (1999). Monitoring, moral hazard, and market power: a model of bank lending. Finance and Economics Discussion Series 37, Board of Governors of the Federal Reserve System.
- Degryse, H., and S. Ongena. 2001. Bank relationships and firm profitability. *Financial Management* 30: 9–34.
- Dennis, S.A., and D.J. Mullineaux. 2000. Syndicated loans. *Journal of Financial Intermediation* 9: 404–426.
- Dennis, S., D. Nandy, and I.G. Sharpe. 2000. The determinants of contract terms in bank revolving credit agreements. *Journal of Financial and Quantitative Analysis* 35: 87–110.
- Detragiache, E., P.G. Garella, and L. Guiso. 2000. Multiple versus single banking relationships, theory and evidence. *Journal of Finance* 55: 1133–1161.
- Doyle, J.T. (2003). Credit risk measurement and pricing in performance pricing-based debt contracts. Working Paper, University of Michigan.
- Esty, B., and W. Megginson. 2003. Creditor rights, enforcement, and debt ownership structure: evidence from the global syndicated loan market. *Journal of Financial and Quantitative Analysis* 38 (1): 37–59.
- Flannery, M. 1986. Asymmetric information and risky debt maturity choice. *Journal of Finance* 41: 18–38.
- Gottesman, A.A., and G.S. Roberts. 2004. Maturity and corporate loan pricing. *Financial Review* 38: 55–77.
- Gottesman, A.A. and Roberts, G.S. (2005). “Loan rates and collateral.” Working Paper, York University.
- Guedes, J., and T. Opler. 1996. The determinants of the maturity of new corporate debt issues. *Journal of Finance* 51: 1809–1833.
- Hannan, T.H. 1997. Market share inequality, the number of competitors, and the HHI: an examination of bank pricing. *Review of Industrial Organization* 12: 23–35.
- Harhoff, D., and T. Korting. 1998. Lending relationships in Germany: empirical evidence from survey data. *Journal of Banking and Finance* 22: 1317–1353.
- Helwege, J., and C.M. Turner. 1999. The slope of the credit yield curve for speculative-grade issuers. *Journal of Finance* 54: 1869–1884.
- Hester, D.D. 1979. Customer relationships and terms of loans: evidence from a pilot survey. *Journal of Money, Credit, and Banking* 11: 349–357.
- John, K., A.W. Lynch, and M. Puri. 2003. Credit ratings, collateral and loan characteristics: implications for yield. *Journal of Business* 76 (3): 371–410.(July).[0].
- Jones, J., Lang, W., and Nigro, P. (2000). Recent trends in bank loan syndications: evidence for 1995 to 1999. Working Paper, Office of the Controller of the Currency.
- Kale, J., and T.H. Noe. 1990. Risky debt maturity choice in a sequential game equilibrium. *Journal of Financial Research* 13: 155–166.
- Koskela, E. 1983. Credit rationing and non-price loan terms. *Journal of Banking and Finance* 7: 405–416.
- Lee, S.W. and Mullineaux, D.J. (2001). The size and composition of commercial lending syndicates. Working Paper, University of Kentucky.
- Melnik, A., and S. Plaut. 1986. Loan commitment contracts, terms of lending, and credit allocation. *Journal of Finance* 41: 425–435.
- Merton, R.C. 1974. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* 29: 449–470.
- Myers, S. 1977. The determinants of corporate borrowing. *Journal of Financial Economics* 5: 147–176.
- Panyagometh, K., Roberts, G.S., and Gottesman, A.A. (2004). The relation between performance pricing covenants and corporate loan spreads. Working Paper, York University.

- Peterson, M.A., and R.G. Rajan. 1995. The effect of credit market competition on lending relationships. *Quarterly Journal of Economics* 110: 403–444.
- Pozzolo, A.F. (2002). Collateralized lending and borrowers' riskiness. Working Paper, Banca d' Italia.
- Rajan, R.G. 1992. Insiders and outsiders: the choice between informed and arm's-length debt. *Journal of Finance* 47: 1367–1400.
- Saunders, A., and L. Allen. 2002. *Credit risk measurement: New approaches to value at risk and other paradigms*. 2nd ed. New York: Wiley.
- Standard and Poor's. (2004). Corporate defaults in 2003 recede from recent highs. Standard and Poor's, January.
- Strahan, P.E. (1999). Borrower risk and the price and nonprice terms of bank loans. Working Paper, Federal Reserve Bank of New York.
- Treacy, W. F., and Carey, M.S. (1998). Credit risk rating at large US banks. *Federal Reserve Board Bulletin*, (November): 897–921.
- von Thadden, E. L. (1992). The commitment of finance, duplicated monitoring, and the investment horizon. ESF-CEPR Working Paper No. 27 in Financial Markets, 27 London.

Yan He

Abstract

A and B shares exist in the Chinese stock markets. A shareholders are domestic investors and B shareholders are foreign investors. During the early-and mid-1990s, B shares were traded at a discount relative to A shares, and B-share returns were higher than A-share returns. It is found that B-share market has persistent higher bid-ask spreads than the A-share market and traders in the B-share market bear higher informed trading and other transaction costs. In addition, the higher volatility of B-share returns can be attributed to the higher market making costs in the B-share market.

Keywords

Asymmetric information · Bid-ask spread · Chinese A shares · Chinese B shares · Return · Shanghai stock exchange · Shenzhen stock exchange · Stock · Stock ownership · Transaction costs · Volatility

The development of equity markets in China started in early 1990s. Open for business in December 1990, the Shanghai Stock Exchange (SHSE) and the Shenzhen Stock Exchange (SZSE) are the two major securities exchanges in China. By 1998, the SHSE had raised a total of RMB140.814 billion for listed companies and the SZSE had raised a total of RMB 128 billion for listed companies. The two exchanges played an important role in promoting the restructuring of state-owned enterprises.

Stock shares in China are divided into two broad categories: untradable and tradable. By the end of 1998, the total untradable equity of the listed companies was 166.484 billion shares (i.e. 65.89 % of the total equity of the listed companies), allocated as follows: (1) shares owned by government, 86.551 billion; (2) shares

Y. He (✉)

Indiana University Southeast, New Albany, IN, USA

e-mail: yanhe@ius.edu

owned by legal persons, 71.617 billion; (3) shares owned by employees and others, 8.317 billion. Outstanding tradable shares totaled 86.193 billion shares (i.e. 34.11 % of the total equity of the listed companies), allocated as follows: (1) Class A shares, 60.803 billion; (2) Class B shares, 13.395 billion; and (3) Class H shares, 11.995 billion. Class A shares are owned by domestic investors and traded in the domestic markets. Class B shares are owned by foreign investors but traded in the domestic markets. Class H shares are listed on the Hong Kong Stock Exchange.

China has tightly restricted foreign stock ownership throughout the 1990s. The ownership restriction creates two distinct groups of investors: the domestic and foreign investors. Class A shares are domestic shares and class B shares are foreign shares. In 1991, the Shanghai Stock Exchange (SHSE) and Shenzhen Stock Exchange (SZSE) began to offer B shares, providing foreign investors with a legal channel to invest in China's equity markets. B shares are also known as Renminbi Special Shares. B shares are issued in the form of registered shares and they carry a face value denominated in Renminbi. B shares are subscribed and traded in foreign currencies, but they are listed and traded in securities exchanges inside China. The B share market has attracted a considerable amount of foreign investors. The Market provides an additional channel for foreign capital to invest in China.

Since March 2001, China has opened its B-share market—previously reserved for overseas investors—to Chinese individuals with foreign currency deposits. However, the impact of the opening up of the B-share market to Chinese is limited, because that market is small compared to the number of Chinese people and institutions' foreign currency holdings. Despite the rising foreign currency deposits in China, Chinese people who have foreign currency holdings still account for a very small proportion of investors.

Tables 1, 2, and 3 are obtained from the China Securities Regulatory Committee.

During the early- and mid-1990s, B shares were traded at a discount relative to A shares, and B-share returns were higher than A-share returns. Su (1999) explains the return premiums on the foreign-owned B shares in the Chinese stock markets by testing a one-period capital asset-price model (CAPM). He concludes that foreign investors are more risk-averse than domestic investors. Sun and Tong (2000) explain the price discount of the B shares by differential demand elasticity. They document that when more H shares and red chips are listed in Hong Kong, the B-share discount becomes larger. In addition, Chui and Kwok (1998) show that the returns on B shares lead the returns on A shares, which induces an asymmetric positive cross-autocorrelation between the returns on B and A shares. They argue that A- and B-share investors have different access to information, and information often reaches the B-share market before it reaches the A-share market.

The Chinese stock markets have grown very rapidly during the late 1990s and early 2000s. A number of studies investigate the return and risk in the newly developed markets. For example, Lee et al. (2001) examine time-series features of stock returns and volatility, as well as the relation between return and volatility in four of China's stock exchanges. On the one hand, test results provide strong evidence of time-varying volatility and show volatility is highly persistent and

Table 1 Trading summary of A and B shares during 11/2000–11/2001

	A or B share listed no. of companies	Total market capitalization (100,000,000 Yuan)	Stock turnover (100,000,000 Yuan)	Stock trading volume (100,000,000)	No. of transactions (10,000)
2000/11	1,063	46,061.78	5,012.27	365.02	5,013
2000/12	1,088	48,090.94	3,737.6	271.35	3,719
2001/01	1,100	48,497.99	3,013.63	220.08	3,082
2001/02	1,110	46,228.75	1,950.05	151.92	2,197
2001/03	1,122	50,908.44	5,095.17	488.33	4,335
2001/04	1,123	51,006.9	5,395.87	422.43	4,720
2001/05	1,129	53,205.49	4,452.16	328.33	3,739
2001/06	1,137	53,630.58	4,917.12	355.5	4,449
2001/07	1,140	46,440.83	3,100.68	228.25	2,983
2001/08	1,151	48,054.63	2,490.85	221.31	2,507
2001/09	1,154	45,831.36	1,766.64	154.67	1,858
2001/10	1,152	43,742.14	1,951.5	181.03	1,914
2001/11	1,153	45,431.59	2,092.26	200.31	2,374

predictable. On the other hand, the results do not show any relation between expected returns and expected risk.

The development in the Chinese markets may affect the risk and return of A- and B-share classes. He and Wu (2003) provide two interesting findings: (1) the daily returns of domestic shares (A shares) and foreign shares (B shares) were almost identical in the late 1990s, while the B-share returns were much higher than the A-share returns during the mid-1990s; (2) the volatility of B-share daily returns was higher than that of A shares, while previous studies have often documented higher return volatility for A shares. (For example, Su and Fleisher (1999) report that A shares have higher volatility than B shares based on the data of mid-1990s).

Since A and B shares are entitled to the same cash flows of a firm and have similar returns, the higher return volatility of B shares is puzzling. The market microstructure theory suggests that both volatility and bid-ask spreads are positively related to asymmetric information (see Kyle, 1985; Easley et al., 1996). According to this theory, higher volatility is caused by higher degree of information asymmetry and participation rate of informed traders in the market, which, in turn, lead to higher trading costs. Thus, the higher volatility of B shares may be due to a more severe asymmetric information problem in the B-share market. If so, we should observe higher trading costs for B shares. Furthermore, Easley et al. (1996) show that spreads and volatility are negatively related to liquidity. Since the order processing cost is the cost of providing liquidity and immediacy, lower liquidity results in higher order processing cost and higher volatility. A recent study by Green et al. (2000) on the London Stock Exchange shows that changes in transaction costs have a significant effect on share price volatility. Moreover, Chordia et al. (2002) document that return volatility is significantly related to quoted spreads. These findings confirm the theoretical prediction that volatility and trading costs are positively correlated.

Table 2 A and B shares offering (1987–1998)

	87	88	89	90	91	92	93	94	95	96	97	98	Total
Shares issued (100MM)	10	25	7	4	5	21	96	91	32	86	268	102	746
A share	10	25	7	4	5	10	43	11	5	38	106	79	343
H share							40	70	15	32	137	13	307
B share						11	13	10	11	16	25	10	96
Capital raised (RMB 100MM)	10	25	7	4		94	375	327	150	425	1,294	837	3,553
A share	10	25	7	4		50	195	50	23	224	655	440	1,687
H share							61	89	31	84	360	38	763
B share						44	38	38	33	47	81	26	307
Rights offering of A and B shares							82	50	63	70	198	335	797

Table 3 Number of listed companies (1990–1998)

Companies	1990	1991	1992	1993	1994	1995	1996	1997	1998
Issuing A share	10	14	35	140	227	242	431	627	727
Issuing B share	0	0	0	6	4	12	16	25	26
Issuing A and B shares	0	0	18	34	54	58	69	76	80
Issuing A and H shares	0	0	0	3	6	11	14	17	18
Total	10	14	53	183	291	323	530	745	851

Therefore, the higher volatility in the B-share market may reflect higher idiosyncratic risk (rather than higher systematic risk) of B-share stocks. The trading risk associated with asymmetric information can be diversified away and therefore it is not systematic risk (see Chordia et al. 2002). Asset-pricing models (e.g. CAPM and APT) suggest that expected returns should be determined by systematic risk. Since higher volatility does not necessarily imply higher systematic risk, it may not be accompanied with higher returns. Su (1999) finds that market risk (measured by market betas) can explain returns of A and B shares, but nonmarket risk variables, such as the variance of returns and firm size, do not systematically affect returns. Thus, the difference in return volatility between the A- and B-share markets may be caused by the difference in idiosyncratic risk. Trading cost, which reflects asymmetric information and liquidity of trading, may explain the B-share market anomaly. For example, if B-share investors incur higher trading costs than A-share investors, the return volatility of B shares would be higher than that of A shares, other things

being equal. In line with the above arguments, He et al. (2003) examine whether the difference in trading costs (or market making costs) between the Chinese A and B shares can explain the difference in return volatility between the two classes of shares. They estimate the end-of-day bid-ask spread and its informed trading and noninformed trading cost components for each stock using daily data in the late 1990s. Their results show that the B-share market has persistent higher bid-ask spreads than the A-share market, and traders in the B-share market bear higher informed trading and other transaction costs. Furthermore, they find that the higher volatility of B-share returns can be attributed to the higher market making costs in the B-share market.

References

- Chordia, T., R. Roll, and A. Subrahmanyam. 2002. Order imbalance, liquidity, and market returns. *Journal of Financial Economics* 65: 111–131.
- Chui, A., and C. Kwok. 1998. Cross-autocorrelation between A shares and B shares in the Chinese Stock Market. *Journal of Financial Research* 21: 333–354.
- Easley, D., N. Kiefer, M. O'Hara, and J. Paperman. 1996. Liquidity, information, and infrequently traded stocks. *Journal of Finance* 51: 1405–1436.
- Green, C.J., P. Maggioni, and V. Murinde. 2000. Regulatory lessons for emerging stock markets from a century of evidence on transactions costs and share price volatility in the London Stock Exchange. *Journal of Banking and Finance* 24: 577–601.
- He, Y., C. Wu, and Y.-M. Chen. 2003. An explanation of the volatility disparity between the domestic and foreign shares in the Chinese Stock Markets. *International Review of Economics and Finance* 12: 171–186.
- Kyle, A. 1985. Continuous auctions and insider trading. *Econometrica* 53: 1315–1335.
- Lee, C.F., G. Chen, and O.M. Rui. 2001. Stock returns and volatility on China's stock markets. *The Journal of Financial Research* 24: 523–544.
- Su, D. 1999. Ownership restrictions and stock prices: evidence from Chinese markets. *Financial Review* 34: 37–56.
- Su, D., and B.M. Fleisher. 1999. Why does return volatility differ in Chinese stock markets? *Pacific-Basin Finance Journal* 7: 557–586.
- Sun, Q., and W. Tong. 2000. The effect of market segmentation on stock prices: the China syndrome. *Journal of Banking and Finance* 24: 1875–1902.

Yan He

Abstract

All NYSE-listed stocks were switched from a fractional to a decimal trading system on January 29, 2001 and all NASDAQ stocks followed suit on April 9, 2001. The conversion to decimal trading in the U.S. markets has significantly reduced bid–ask spreads. This decline is primarily due to the drop in market makers’ costs for supplying liquidity. In addition, rounding becomes less salient after the decimalization. The decrease in bid–ask spreads can be ascribed to the decrease in price rounding, when controlling for the changes in trading variables.

Keywords

Bid–ask spread · Clustering · Decimalization · Decimal trading · Fractional trading · NASDAQ · NYSE · Price improvement · Rounding · Volatility

The minimum increment of trading prices varies substantially with market and location. For instance, pricing of stock, bond, and options markets in the U.S. and Canada had traditionally been denominated in eighths, while in European and Asian markets decimal prices are more common. During the later half of 1990s, the U.S. and Canadian markets underwent substantial changes. Canadian stocks switched from fractions to decimals in April 1996. In the U.S. markets, the minimum tick size was reduced from one-eighth of a dollar to one-sixteenth of a dollar in June 1997. At the beginning of year 2000, the U.S. equity markets were the only major financial markets in the world that traded in fractional increments. This fractional trading practice puts U.S. markets at a competitive disadvantage with foreign markets trading the same securities. In addition, individual investors may have a difficulty in determining the differences between increasingly smaller fractions.

Y. He (✉)

Indiana University Southeast, New Albany, IN, USA

e-mail: yanhe@ius.edu

To make the U.S. securities markets more competitive globally and their prices easier to decipher, the Securities Industry Association and the Securities and Exchange Commission decided to convert the U.S. equity and exchange-traded options markets from fractional to decimal trading. The NYSE selected seven pilot securities for a decimal pricing test on August 28, 2000, another 57 securities were added to the pilot program on September 25, 2000, and another 94 were added on December 4, 2000. The NASDAQ market began its decimal test with 14 securities on March 12, 2001, and another 197 securities were added on March 26, 2001. All NYSE-listed stocks were switched to a decimal trading system on January 29, 2001 and all NASDAQ stocks followed suit on April 9, 2001.

Recently, a number of studies have generated interesting findings about the effects of decimalization on return volatility and bid–ask spreads. They report that decimalization affects bid–ask spreads, volatility, quote size, and price improvement frequency (or the probability of trades within the quoted bid–ask spreads). First of all, it was shown that the recent conversion to decimal trading in the U.S. markets has significantly reduced bid–ask spreads (see NYSE 2001; NASDAQ 2001; Chakravarty et al. 2001a, b; Chung et al. 2001; Gibson et al. 2002). These findings coincide with two earlier studies (Ahn et al. 1998; Bacidore 1997) on the Toronto Stock Exchange (TSE). Bacidore et al. (2001b) examine a wide range of market quality issues on the NYSE post-decimalization, and find that an increase in the aggressiveness of limit order pricing results in narrower bid–ask spreads. Chung et al. (2004) examine the relationship between NASDAQ trading costs and the extent of order preferencing. They document lower order preferencing and a positive relationship between the bid–ask spread and the proportion of internalized volume on NASDAQ after decimalization. Second, Bessembinder (2003) and NASDAQ (2001) show that intraday return volatility has declined, and there is no evidence of systematic reversals in quotation changes. Thus, it appears that the NYSE and NASDAQ markets are able to supply sufficient liquidity in the wake of decimalization. Third, Bessembinder (2003) presents that quote size decreases after decimalization. Jones and Lipson (2001) and Goldstein and Kavajecz (2000) report decreases in limit-order book depth after an earlier NYSE tick size reduction, and Bacidore et al. (2001a) report decreases in limit-order book depth after the decimalization on the NYSE. Finally, Bacidore et al. (2001b) and Bessembinder (2003) find evidence that the percentage of orders experiencing price improvement (i.e. executed within the quotes) increases on the NYSE after decimalization, though the dollar amount of price improvement falls. According to Coughenour and Harris (2003), decimal trading effectively relaxes the public order precedence rule and gives specialists more price points within the bid–ask spread on which to quote aggressively. This allows specialist trading firms of all size to trade more often inside the current quote and so the probability that a trade occurs inside the quotes becomes higher.

Almost all the above studies document the changes in “total” return volatility and spreads of transactions. He and Wu (2004) examine the composition of return volatility, serial correlation, and trading costs before and after the decimalization on the NYSE. Specifically, they decompose the variance of price changes into components associated with public news, rounding errors, and market-making

frictions (asymmetric information and liquidity costs). First, the test results show that both variance components due to market-making frictions (or bid–ask spreads) and rounding errors decline considerably after decimalization, while the variance component due to public news shocks remains unchanged. Second, the serial correlation of price changes is significantly reduced after decimalization, indicating a weakened bid–ask bounce effect as a result of decimal trading. Finally, bid–ask spreads decline substantially after decimalization and this decline is primarily due to the drop in market makers' costs for supplying liquidity.

In addition to volatility and transaction costs, the recent decimalization also provides an opportunity to revisit the issue of price rounding. Since traders often choose to use a larger price increment than the minimum tick, prices tend to cluster on certain fractions or decimals even when the tick is small. (See Ball et al. (1985) for gold trading; Brown et al. (1991) for silver; Goodhart and Curcio (1992) for foreign exchange; and Aitken, et al. (1995) for Australian stocks.) Harris (1999) predicts that the conversion to decimal trading would lead to lower execution costs. Bessembinder (2003) shows that bid–ask spreads have declined after the decimalization.

He and Wu (2003) investigate the pattern of price rounding before and after decimal trading and its effect on bid–ask spreads for NYSE stocks by using the second pilot sample which includes 57 NYSE securities. Prior to September 25, 2000, these stocks were traded on sixteenths. Since then, they have been traded on pennies. First, since decimal trading leads to a finer price grid or a set of less discrete prices, it is expected to observe a decline in frequencies of rounding on integers, halves, and quarters. Second, although frequencies of rounding on integers, halves, and quarters may decline after decimalization, it is expected that cross-sectionally the relationship between rounding and trading variables and the relationship between execution costs and rounding will stay the same. That is, the sensitivity of trading variables to rounding and the sensitivity of execution costs to rounding should remain unchanged because the fundamentals of the market do not change as a result of decimalization. Finally, consistent with the arguments of Harris (1997, 1999), it is expected to find a significant relationship between the decrease in execution costs and the decrease in rounding after decimalization, when controlling for the changes in stock features. If fractional pricing indeed allows market makers to keep bid–ask spreads artificially high to earn a positive rent, a conversion to decimal trading should reduce price rounding, decrease market makers' rents, and cause a fall in bid–ask spreads.

The empirical results of He and Wu (2003) show that although rounding is pervasive in transaction prices, bids, and asks in both the pre- and post-decimalization periods, it has become less salient after the decimalization. The cross-sectional relationship between rounding and trading variables is similar before and after the decimalization, and so is the relationship between execution costs and rounding when trading variables are held constant for each stock. More importantly, the quoted and effective bid–ask spreads decrease after the decimalization, and this decrease can be ascribed to the decrease in price rounding when controlling for the changes in trading variables.

References

- Ahn, H.-J., Q.C. Cao, and H. Choe. 1998. Decimalization and competition among stock markets: Evidence from the Toronto stock exchange cross-listed securities. *Journal of Financial Markets* 1: 51–87.
- Aitken, M., Brown, P., Buckland, C., Izan, H.Y., and Walter, T. (1995). Price clustering on the Australian stock exchange. Working Paper, University of Western Australia.
- Bacidore, J. 1997. The impact of decimalization on market quality: an empirical investigation of the Toronto Stock Exchange. *Journal of Financial Inter-mediation* 6: 92–120.
- Bacidore, J., Battalio, R., and Jennings, R. (2001a). Order submission strategies, liquidity supply, and trading in pennies on the New York Stock Exchange. Working Paper, Indiana University.
- Bacidore, J., Battalio, R., Jennings, R., and Farkas, S. (2001b). Changes in order characteristics, displayed liquidity, and execution quality on the NYSE around the switch to decimal pricing. Working Paper, The New York Stock Exchange.
- Ball, C.A., W.N. Torous, and A.E. Tshoeogl. 1985. The degree of price resolution: the case of the gold market. *Journal of Futures Markets* 5: 29–43.
- Bessembinder, H. 2003. Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis* 13: 19–42.
- Brown, S., P. Laux, and B. Schachter. 1991. On the existence of an optimal tick size. *Review of Futures Markets* 10: 50–72.
- Chakravarty, S., Harris, S., and Wood, R. (2001a). Decimal trading and market impact. Working Paper, University of Memphis.
- . (2001b). Decimal trading and market impact: The Nasdaq experience. Working Paper, University of Memphis.
- Chung, K., Van-Ness, B., and Van-Ness, R. (2001). Are Nasdaq stocks more costly to trade than NYSE stocks? Evidence after decimalization. Working Paper, Kansas State University.
- Chung, K., C. Chuwonganant, and T. McCormick. 2004. Order preferencing and market quality on Nasdaq before and after decimalization. *Journal of Financial Economics* 71: 581–612.
- Coughenour, J. and Harris, L. (2003). Specialist profits and the minimum price increment. Working Paper.
- Gibson, S., Singh, R., and Yerramilli, V. (2002). The effect of decimalization on the components of the bid-ask spreads. Working Paper, Cornell University.
- Goldstein, M., and K. Kavajecz. 2000. Eighths, sixteenths and market depth: Changes in tick size and liquidity provision on the NYSE. *Journal of Financial Economics* 56: 125–149.
- Goodhart, C. and Curcio, R. (1992). Asset price discovery and price clustering in the foreign exchange market. Working Paper, London School of Business.
- Harris, L. (1997). Decimalization: a review of the arguments and evidence. Working Paper, University of Southern California.
- . (1999). Trading in pennies: A survey of the issues. Working Paper, University of Southern California.
- He, Y. and Wu, C. (2003). The effects of decimalization on return volatility components, serial correlation, and trading costs. Working Paper.
- He, Y., and C. Wu. 2004. Price rounding and bid-ask spreads before and after the decimalization. *International Review of Economics and Finance* 13: 19–42.
- Jones, C., and M. Lipson. 2001. Sixteenths: direct evidence on institutional trading costs. *Journal of Financial Economics* 59: 253–278.
- NASDAQ. (2001). “The impact of decimalization on the NASDAQ stock market,” prepared by NASDAQ Research Department.
- NYSE. (2001). “Comparing bid-ask spreads on the NYSE and NASDAQ immediately following NASDAQ decimalization,” prepared by NYSE Research Department.

Yan He

Abstract

Several important trading rules were introduced in NASDAQ in 1997. The trading reforms have significantly reduced bid–ask spreads on NASDAQ. This decrease is due to a decrease in market-making costs and/or an increase in market competition for order flows. In addition, in the post-reform period, the spread difference between NASDAQ and the NYSE becomes insignificant with the effect of informed trading costs controlled.

Keywords

Bid–ask spread · Informed trading costs · NASDAQ · NYSE · Reforms · SEC · SEC order handling rules · The actual size rule · The sixteenths minimum increment rule · Trading rules

The National Association of Securities Dealers (NASD) was established in 1939. Its primary role was to regulate the conduct of the over-the-counter (OTC) segment of the securities industry. In the middle of 1960s, the NASD developed an electronic quote dissemination system, and in 1971, the system began formal operation as the National Association of Securities Dealers Automated Quotations (NASDAQ) system. By the mid-1980s, timely last-sale price and volume information were made available on the terminals. Through the late 1980s and the early 1990s, more functions were added to the system. For instance, the Small Order Execution System (SOES) was introduced in 1988, and the Electronic Communication Networks (ECN) was introduced in the 1990s. Services provided by the NASDAQ network include quote dissemination, order routing, automatic order execution, trade reporting, last sale, and other general market information.

Y. He (✉)

Indiana University Southeast, New Albany, IN, USA

e-mail: yanhe@ius.edu

NASDAQ is a dealer market, and it is mainly quote driven. On NASDAQ, the bid–ask quotes of competing dealers are electronically disseminated to brokers' offices, and the brokers send the customer order flow to the dealers who have the best quotes. In comparison, the New York Stock Exchange (NYSE) is an auction market, and it is mainly order driven.

Several important trading rules were introduced in NASDAQ in 1997, including the SEC Order Handling Rules, the Sixteenths Minimum Increment Rule, and the Actual Size Rule. The experimentation of the new rules started on January 20, 1997. The SEC Order Handling Rules were applied to all the NASDAQ stocks in October 1997. The Actual Size Rule was applied to 50 NASDAQ stocks on January 20, 1997 and 104 additional stocks on November 10, 1997. The Sixteenths Minimum Increment Rule was applied to all the stocks in NASDAQ on June 2, 1997. The following table provides a detailed implementation schedule for the new trading rules.

NASDAQ implemented the Order Handling Rules according to a phased-in schedule. On January 20, 1997, the first group of 50 stocks became subject to the Order Handling Rules. The SEC Order Handling Rules include the Limit Order Display Rule, the ECN Rule, and the Relaxation of the Excess Spread Rule (Table 1).

The Limit Order Display Rule requires displaying customer limit orders that are priced better than a market maker's quote, or adding them to the size associated with a market maker's quote when it is the best price in the market. Before the new trading rules, limit orders on NASDAQ were only offered to the market makers. The Limit Order Display Rule promotes and facilitates the public availability of quotation information, fair competition, market efficiency, the best execution of

Table 1 New trading rules' implementation schedule

Date	Number of stocks affected by the rules	Rules implemented
01/20/1997	50 NASDAQ stocks	The SEC Order Handling Rules
		The Actual Size Rule
	The same 50 NASDAQ stocks	The Relaxation of the Excess Spread Rule
	All the NASDAQ stocks	
02/10/1998	51 NASDAQ stocks added	The SEC Order Handling Rules
02/24/1997	52 NASDAQ stocks added	The SEC Order Handling Rules
04/21/1997–07/07/1997	563 NASDAQ stocks added	The SEC Order Handling Rules
06/02/1997	All NASDAQ stocks with bid price not less than \$10	The Sixteenths Minimum Increment Rule
08/04/1997	250 NASDAQ stocks added	The SEC Order Handling Rules
08/11/1997	251 NASDAQ stocks added	The SEC Order Handling Rules
09/08/1997–10/13/1997	800 NASDAQ stocks/week added	The SEC Order Handling Rules
10/13/1997	All NASDAQ stocks	The SEC Order Handling Rules
11/10/1997	104 stocks added	The Actual Size Rule

customer orders, and the opportunity for investors' orders to be executed without the participation of a dealer. By virtue of the Limit Order Display Rule, investors now have the ability to directly advertise their trading interests to the marketplace, thereby allowing them to compete with market maker quotations, and affect bid-ask spreads.

The ECN Rule requires market makers to display in their quotes any better-priced orders that the market maker places into an ECN. The ECN Rule was implemented partially because market participants had increasingly been using ECNs to display different prices to different market participants. In particular, NASDAQ was concerned that the reliability and completeness of publicly available quotations were compromised because market makers could widely disseminate prices through ECNs superior to the quotation information they disseminate on a general basis through NASDAQ. Accordingly, the ECN Rule was adopted to require the public display of such better-priced orders.

Prior to January 20, 1997, NASDAQ continuously calculated for each stock the average of the three narrowest individual spreads among all dealers' spreads. The Excess Spread Rule (ESR) forced all dealers to keep their spreads within 125% of this average. On January 20, 1997, the ESR was amended for all NASDAQ stocks to stipulate that each dealer's average spread during the month could not exceed 150% of the three lowest average spreads over the month. The new ESR defines compliance on a monthly basis rather than continuously, placing no limits on the market makers' ability to vary their spreads during the month as long as their monthly average is in compliance.

The Actual Size Rule is a by-product of the Order Handling Rules. This rule repeals the regulatory minimum quote size (1,000 shares). With the implementation of the SEC's Order Handling Rules, the 1,000 share minimum quote size requirements impose unnecessary regulatory burdens on market makers. Since the investors are allowed to display their own orders on NASDAQ according to the Limit Order Display Rule, the regulatory justification for the 1,000 share minimum quote size requirements is eliminated. So, it is appropriate to treat NASDAQ market makers in a manner equivalent to exchange specialists, and not subject them to the 1,000 share minimum quote size requirements. On January 20, 1997, 50 pilot stocks became subject to the Actual Size Rule. These 50 stocks also became subject to the SEC Order Handling Rules. On November 10, 1997, the pilot program was expanded to an additional 104 stocks. After 1997, the Rule was implemented to all stocks on NASDAQ.

The Sixteenths Minimum Increment Rule requires that the minimum quotation increment be reduced from one-eighth to one-sixteenth of a dollar for all securities with a bid price of \$10 or higher. On June 2, 1997, NASDAQ reduced the minimum quotation increment from one-eighth to one-sixteenth of a dollar for all NASDAQ securities with a bid price of \$10 or higher. The reduction is expected to tighten quoted spreads and enhance quote competition. Furthermore, it complements the Order Handling Rules by allowing orders to be displayed in increments finer than one-eighth of a dollar. Specifically, the opportunity is increasing for small customers and ECN limit orders to drive the inside market.

Overall, all these new rules were designed to enhance the quality of published quotation, promote competition among dealers, improve price discovery, and increase liquidity. Under these rules, NASDAQ is transformed from a pure quote driven market to a more order driven market. Successful implementation of these rules should result in lower bid–ask spreads by either reducing order execution costs or dealers’ profits.

Before 1997, a host of studies compared trading costs between NASDAQ and the NYSE based on the old trading rules. It is documented that bid–ask spreads or execution costs are significantly higher on NASDAQ than on the NYSE. Researchers debate whether NASDAQ bid–ask spreads are competitive enough to reflect market-making costs. Christie and Schultz (1994) find that NASDAQ dealers avoid odd-eighth quotes. This evidence is interpreted as consistent with tacit collusion, due to which bid–ask spreads are inflated above the competitive level. Moreover, Huang and Stoll (1996) and Bessembinder and Kaufman (1997) contend that higher spreads on NASDAQ cannot be attributed to informed trading costs.

Since the Securities and Exchange Committee (SEC) changed some important trading rules on NASDAQ in 1997, studies attempt to assess the effect of these reforms on market performance. Barclay et al. (1999) report that the reforms have significantly reduced bid–ask spreads on NASDAQ. Bessembinder (1999) finds that trading costs are still higher on NASDAQ than on the NYSE even after NASDAQ implemented new trading rules. Weston (2000) shows that the informed trading and inventory costs on NASDAQ remain unchanged after the reforms, and that the reforms have primarily reduced dealers’ rents and improved competition among dealers on NASDAQ. He and Wu (2003a) report further evidence of the difference in execution costs between NASDAQ and the NYSE before and after the 1997 market reforms. In the prereform period the NASDAQ–NYSE disparity in bid–ask spreads could not be completely attributed to the difference in informed trading costs. However, in the postreform period the spread difference between these two markets becomes insignificant with the effect of informed trading costs controlled. In addition, He and Wu (2003b) examine whether the decrease in bid–ask spreads on NASDAQ after the 1997 reforms is due to a decrease in market-making costs and/or an increase in market competition for order flows. Their empirical results show that lower market-making costs and higher competition significantly reduce bid–ask spreads.

References

- Barclay, M.J., W.G. Christie, J.H. Harris, E. Kandel, and P.H. Schultz. 1999. Effects of market reform on the trading costs and depths of NASDAQ stocks. *Journal of Finance* 54: 1–34.
- Bessembinder, H. 1999. Trade execution costs on NASDAQ and the NYSE: A post-reform comparison. *Journal of Financial and Quantitative Analysis* 34: 387–407.
- Bessembinder, H., and H. Kaufman. 1997. A comparison of trade execution costs for NYSE and NASDAQ-listed stocks. *Journal of Financial and Quantitative Analysis* 32: 287–310.

- Christie, W.G., and P.H. Schultz. 1994. Why do NASDAQ market makers avoid odd-eighth quotes? *Journal of Finance* 49: 1813–1840.
- He, Y., and C. Wu. 2003a. The post-reform bid-ask spread disparity between NASDAQ and the NYSE. *Journal of Financial Research* 26: 207–224.
- . 2003b. What explains the bid-ask spread decline after NASDAQ reforms? *Financial Markets, Institutions & Instruments* 12: 347–376.
- Huang, R.D., and H.R. Stoll. 1996. Dealer versus auction markets: a paired comparison of execution costs on NASDAQ and the NYSE. *Journal of Financial Economics* 41: 313–357.
- Weston, J. 2000. Competition on the NASDAQ and the impact of recent market reforms. *Journal of Finance* 55: 2565–2598.



Randall A. Heron and Wilbur G. Lewellen

Abstract

Under the state corporate chartering system in the U.S., managers may seek shareholder approval to reincorporate the firm in a new state, regardless of the firm's physical location, whenever they perceive that the corporate legal environment in the new state is better for the firm. Legal scholars continue to debate the merits of this system, with some arguing that it promotes contractual efficiency and others arguing that it often results in managerial entrenchment. We discuss the contrasting viewpoints on reincorporations and then summarize extant empirical evidence on why firms reincorporate, when they reincorporate, and where they reincorporate to. We conclude by discussing how the motives managers offer for reincorporations, and the actions they take upon reincorporating, influence how stock prices react to reincorporation decisions.

Keywords

Antitake-over · Contractual efficiency · Corporate charter · Corporate law · Delaware · Director liability · Incorporation · Managerial entrenchment · Reincorporation · Shareholders · Takeover defenses

R. A. Heron (✉)
Indiana University, Indianapolis, IN, USA
e-mail: rheron@iupui.edu

W. G. Lewellen
Purdue University, Indianapolis, IN, USA

19.1 Introduction

Modern corporations have been described as a “nexus of contractual relationships” that unites the providers and users of capital in a manner that is superior to alternative organizational forms. While agency costs are an inevitable consequence of the separation of ownership and control that characterizes corporations, the existence of clearly specified contractual relationships serves to minimize those costs. As Jensen and Meckling (1976, p. 357) noted:

The publicly held business corporation is an awesome social invention. Millions of individuals voluntarily entrust billions of dollars, francs, pesos, etc., of personal wealth to the care of managers on the basis of a complex set of contracting relationships which delineate the rights of the parties involved. The growth in the use of the corporate form as well as the growth in market value of established corporations suggests that, at least up to the present, creditors and investors have by and large not been disappointed with the results, despite the agency costs inherent in the corporate form.

Agency costs are as real as any other costs. The level of agency costs depends among other things on statutory and common law and human ingenuity in devising contracts. Both the law and the sophistication of contracts relevant to the modern corporation are the products of a historical process in which there were strong incentives for individuals to minimize agency costs. Moreover, there were alternative organizational forms available, and opportunities to invent new ones. Whatever its shortcomings, the corporation has thus far survived the market test against potential alternatives.

Under the state corporate chartering system that prevails in the U.S., corporate managers can affect the contractual relationships that govern their organizations through the choice of a firm’s state of incorporation. Each state has its own distinctive corporate laws and established court precedents that apply to firms incorporated in the state. Thus, corporations effectively have a menu of choices for the firm’s legal domicile, from which they may select the one they believe is best for their firm and/or themselves. The choice is not constrained by the physical location either of the firm’s corporate headquarters or its operations. A firm whose headquarters is in Texas may choose Illinois to be its legal domicile, and vice versa. Corporations pay fees to their chartering states, and these fees vary significantly across states, ranging up to \$150,000 annually for large companies incorporated in Delaware. State laws of course evolve over time, and managers may change their firm’s legal domicile – subject to shareholder approval – if they decide the rules in a new jurisdiction would be better suited to the firm’s changing circumstances. This is the process referred to as *reincorporation*, and it is our topic of discussion here.

19.2 Competition Among States for Corporate Charters

There has been a long-running debate among legal and financial scholars regarding the pros and cons of competition among states for corporate charters. Generally speaking, the proponents of competition claim that it gives rise to a wide variety of contractual relationships across states, which allows the firm to choose the legal

domicile that serves to minimize its organizational costs and thereby maximize its value. This “Contractual Efficiency” viewpoint, put forth by Dodd and Leftwich (1980), Easterbrook and Fischel (1983), Baysinger and Butler (1985), and Romano (1985), implies the existence of a determinate relationship between a company’s attributes and its choice of legal residency. Such attributes may include: (1) the nature of the firm’s operations, (2) its ownership structure, and (3) its size. The hypothesis following from this viewpoint is that firms that decide to reincorporate do so when the firm’s characteristics are such that a change in legal jurisdiction increases shareholder wealth by lowering the collection of legal, transactional, and capital-market-related costs it incurs.

Other scholars, however, argue that agency conflicts play a significant role in the decision to reincorporate, and that these conflicts are exacerbated by the competition among states for the revenues generated by corporate charters and the economic side effects that may accompany chartering (e.g., fees earned in the state for legal services). This position, first enunciated by Cary (1974), is referred to as the “Race-to-the-Bottom” phenomenon in the market for corporate charters. The crux of the Race-to-the-Bottom argument is that states that wish to compete for corporate chartering revenues will have to do so along dimensions that appeal to corporate management.

Hence, states will allegedly distinguish themselves by tailoring their corporate laws to serve the self-interest of managers at the expense of corporate shareholders. This process could involve creating a variety of legal provisions that would enable management to increase its control of the corporation, and thus to minimize the threats posed by outside sources. Examples of the latter would include shareholder groups seeking to influence company policies, the threat of holding managers personally liable for ill-advised corporate decisions, and – perhaps most important of all – the threat of displacement by an alternative management team. These threats, considered by many to be necessary elements in an effective system of corporate governance, can impose substantial personal costs on senior managers. That may cause managers to act in ways consistent with protecting their own interests – through job preservation and corporate risk reduction – rather than serving the interests of shareholders. If so, competition in the market for corporate charters will diminish shareholder wealth as states adopt laws that place restrictions on the disciplinary force of the market for corporate control (see Bebchuk 1992; Bebchuk and Ferrell 1999; Bebchuk and Cohen 2003).

Here, we examine the research done on reincorporation and discuss the support that exists for the contrasting views of both the Contractual Efficiency and Race-to-the-Bottom proponents. In the process, we shall highlight the various factors that appear to play an influential role in the corporate chartering decision.

19.3 Why, When, and Where to Reincorporate

To begin to understand reincorporation decisions, it is useful to review the theory that relates a firm’s choice of chartering jurisdiction to the firm’s attributes, the evidence as to what managers *say* when they propose reincorporations to their shareholders, and what managers *actually do* when they reincorporate their firms.

Central to the Contractual Efficiency view of competition in the market for corporate charters is the notion that the optimal chartering jurisdiction is a function of the firm's attributes. Reincorporation decisions therefore should be driven by changes in a firm's attributes that make the new state of incorporation a more cost-effective legal jurisdiction. Baysinger and Butler (1985) and Romano (1985) provide perhaps the most convincing arguments for this view.

Baysinger and Butler theorize that the choice of a strict vs. a liberal incorporation jurisdiction depends on the nature of a firm's ownership structure. The contention is that states with strict corporate laws (i.e., those that provide strong protections for shareholder rights) are better suited for firms with concentrated share ownership, whereas liberal jurisdictions promote efficiency when ownership is widely dispersed. According to this theory, holders of large blocks of common shares will prefer the pro-shareholder laws of strict states, since these give shareholders the explicit legal remedies needed to make themselves heard by management and allow them actively to influence corporate affairs. Thus, firms chartered in strict states are likely to remain there until ownership concentration decreases to the point that legal controls may be replaced by market-based governance mechanisms.

Baysinger and Butler test their hypothesis by comparing several measures of ownership concentration in a matched sample of 302 manufacturing firms, half of whom were incorporated in several strict states (California, Illinois, New York, and Texas) while the other half had reincorporated out of these states. In support of their hypothesis, Baysinger and Butler found that the firms that stayed in the strict jurisdictions exhibited significantly higher proportions of voting stock held by major blockholders than was true of the matched firms who elected to reincorporate elsewhere. Importantly, there were no differences between the two groups in financial performance that could explain why some left and others did not. Collectively, the results were interpreted as evidence that the corporate chartering decision is affected by ownership structure rather than by firm performance.

Romano (1985) arrived at a similar conclusion from what she refers to as a "transaction explanation" for reincorporation. Romano suggests that firms change their state of incorporation "at the same time they undertake, or anticipate engaging in, discrete transactions involving changes in firm operation and/or organization" (p. 226). In this view, firms alter their legal domiciles at key times to destination states where the laws allow new corporate policies or activities to be pursued in a more cost-efficient manner. Romano suggests that, due to the expertise of Delaware's judicial system and its well-established body of corporate law, the state is the most favored destination when companies anticipate legal impediments in their existing jurisdictions. As evidence, she cites the high frequency of reincorporations to Delaware coinciding with specific corporate events such as initial public offerings (IPOs), mergers and acquisitions, and the adoption of antitakeover measures.

In their research on reincorporations, Heron and Lewellen (1998) also discovered that a substantial portion (45%) of the firms that reincorporated in the U.S. between 1980 and 1992 did so immediately prior to their IPOs. Clearly, the process of

becoming a public corporation represents a substantial transition in several respects: ownership structure, disclosure requirements, and exposure to the market for corporate control. Accordingly, the easiest time to implement a change in the firm's corporate governance structure to parallel the upcoming change in its ownership structure would logically be just before the company becomes a public corporation, while control is still in the hands of management and other original investors. Other recent studies also report that the majority of firms in their samples who undertook IPOs reincorporated in Delaware in advance of their stock offerings (Daines and Klausner 2001; Field and Karpoff 2002).

Perhaps the best insights into why managers choose to reincorporate their firms come from the proxy statements of publicly traded companies, when the motivations for reincorporation are reported to shareholders. In the process of the reincorporations of U.S. public companies that occurred during the period from 1980 through 1992, six major rationales were proclaimed by management (Heron and Lewellen 1998): (1) takeover defenses; (2) director liability reduction; (3) improved flexibility and predictability of corporate laws; (4) tax and/or franchise fee savings; (5) conforming legal and operating domicile; and (6) facilitating future acquisitions.

A tabulation of the relative frequencies is provided in Fig. 1. As is evident, the two dominant motives offered by management were to create takeover defenses and to reduce directors' legal liability for their decisions. In addition, managers often cited multiple reasons for reincorporation. The mean number of stated motives was 1.6 and the median was 2. In instances where multiple motives were offered, each is counted once in the compilation in Fig. 1.

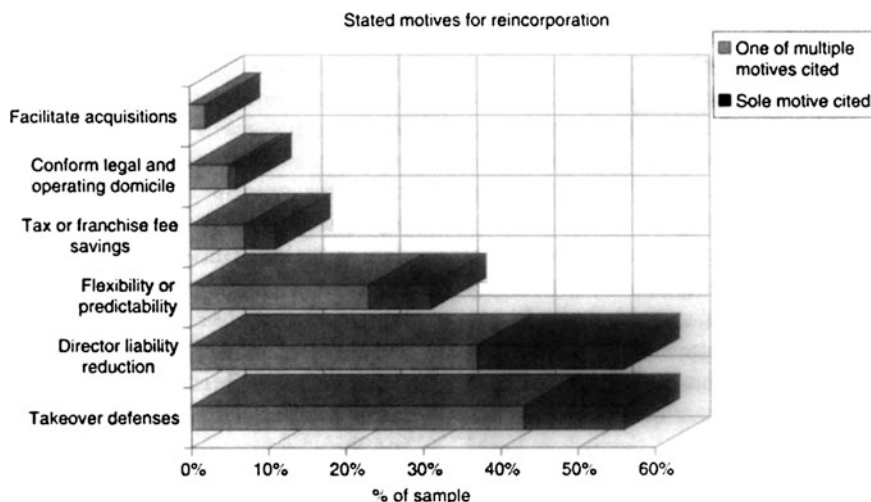


Fig. 1 Stated motives for reincorporation

19.4 What Management Says

It is instructive to consider the stated reincorporation motives in further detail and look at examples of the statements by management that are contained in various proposals, especially those involving the erection of takeover defenses and the reduction of director liability. These, of course, represent provisions that may not be in the best interests of stockholders, as a number of researchers have argued. The other motives listed are both less controversial and more neutral in their likely impact on stockholders, and can be viewed as consistent with Contractual Efficiency arguments for reincorporations. Indeed, reincorporations undertaken for these reasons appear not to give rise to material changes in firms' stock prices (Heron and Lewellen 1998).

19.4.1 Reincorporations That Strengthen Takeover Defenses

Proponents of the Race-to-the-Bottom theory contend that the competition for corporate chartering may be detrimental if states compete by crafting laws that provide managers with excessive protection from the market for corporate control – i.e., from pressures from current owners and possible acquirers to perform their managerial duties so as to maximize shareholder wealth. Although takeover defenses might benefit shareholders if they allow management to negotiate for higher takeover premiums, they harm shareholders if their effect is to entrench poorly performing incumbent managers.

The following excerpts from the proxy statement of Unocal in 1983 provides an example of a proposal to reincorporate for antitakeover reasons:

In addition, incorporation of the proposed holding company under the laws of Delaware will provide an opportunity for inclusion in its certificate of incorporation provisions to discourage efforts to acquire control of Unocal in transactions not approved by its Board of Directors, and for the elimination of shareholder's preemptive rights and the elimination of cumulative voting in the election of directors.

The proposed changes do not result from any present knowledge on the part of the Board of Directors of any proposed tender offer or other attempt to change the control of the Company, and no tender offer or other type of shift of control is presently pending or has occurred within the past two years.

Management believes that attempts to acquire control of corporations such as the Company without approval by the Board may be unfair and/or disadvantageous to the corporation and its shareholders. In management's opinion, disadvantages may include the following:

- nonnegotiated takeover bid may be timed to take advantage of temporarily depressed stock prices;
- nonnegotiated takeover bid may be designed to foreclose or minimize the possibility of more favorable competing bids;

recent nonnegotiated takeover bids have often involved so-called “two-tier” pricing, in which cash is offered for a controlling interest in a company and the remaining shares are acquired in exchange for securities of lesser value. Management believes that “two-tier” pricing tends to stampede shareholders into making hasty decisions and can be seriously unfair to those shareholders whose shares are not purchased in the first stage of the acquisition;

nonnegotiated takeover bids are most frequently fully taxable to shareholders of the acquired corporation.

By contrast, in a transaction subject to approval of the Board of Directors, the Board can and should take account of the underlying and long-term value of assets, the possibilities for alternative transactions on more favorable terms, possible advantages from a tax-free reorganization, anticipated favorable developments in the Company’s business not yet reflected in stock prices, and equality of treatment for all shareholders.

The reincorporation of Unocal into Delaware allowed the firm’s management to add several anti-takeover provisions to Unocal’s corporate charter that were not available under the corporate laws of California, where Unocal was previously incorporated. These provisions included the establishment of a Board of Directors whose terms were staggered (only one-third of the Board elected each year), the elimination of cumulative voting (whereby investors could concentrate their votes on a small number of Directors rather than spread them over the entire slate up for election), and the requirement of a “supermajority” shareholder vote to approve any reorganizations or mergers not approved by at least 75% of the Directors then in office. Two years after its move to Delaware, Unocal was the beneficiary of a court ruling in the *Unocal vs. Mesa* case [493 A.2d 946 (Del. 1985)], in which the Delaware Court upheld Unocal’s discriminatory stock repurchase plan as a legitimate response to Mesa Petroleum’s hostile takeover attempt.

The Unocal case is fairly representative of the broader set of reincorporations that erected takeover defenses. Most included antitakeover charter amendments that were either part of the reincorporation proposal or were made possible by the move to a more liberal jurisdiction and put to a shareholder vote simultaneously with the plan of reincorporation. In fact, 78% of the firms that reincorporated between 1980 and 1992 implemented changes in their corporate charters or other measures that were takeover deterrents (Heron and Lewellen 1998). These included eliminating cumulative voting, initiating staggered Board terms, adopting supermajority voting provisions for mergers, and establishing so-called “poison pill” plans (which allowed the firm to issue new shares to existing stockholders in order to dilute the voting rights of an outsider who was accumulating company stock as part of a takeover attempt).

Additionally, Unocal reincorporated from a strict state known for promoting shareholder rights (California) to a more liberal state (Delaware) whose laws were

more friendly to management. In fact, over half of the firms in the sample studied by Heron and Lewellen (1998), that cited antitakeover motives for their reincorporations, migrated *from* California, and 93% migrated *to* Delaware. A recent study by Bebchuk and Cohen (2003) that investigates how companies choose their state of incorporation reports that strict shareholder-right states that have weak anti-takeover statutes continue to do poorly in attracting firms to charter in their jurisdictions.

Evidence on how stock prices react to reincorporations conducted for anti-takeover reasons suggests that investors perceive them to have a value-reducing management entrenchment effect. Heron and Lewellen (1998) report statistically significant (at the 95% confidence level) abnormal stock returns of -1.69% on and around the dates of the announcement and approval of reincorporations when management cites only antitakeover motives. In the case of firms that actually *gained* additional takeover protection in their reincorporations (either by erecting specific new takeover defenses or by adopting coverage under the anti-takeover laws of the new state of incorporation), the abnormal stock returns averaged a statistically significant -1.62% . For firms whose new takeover protection included poison pill provisions, the average abnormal returns were fully -3.03% and only one-sixth were positive (both figures statistically significant). Taken together with similar findings in other studies, the empirical evidence therefore supports a conclusion that “defensive” reincorporations diminish shareholder wealth.

19.4.2 Reincorporations That Reduce Director Liability

The level of scrutiny placed on directors and officers of public corporations was greatly intensified as a result of the Delaware Supreme Court’s ruling in the 1985 *Smith vs. Van Gorkom* case [488 A.2d 858 (Del. 1985)]. Prior to that case, the Delaware Court had demonstrated its unwillingness to use the benefit of hindsight to question decisions made by corporate directors that turned out after the fact to have been unwise for shareholders. The court provided officers and directors with liability protection under the “business judgment” rule, as long as it could be shown that they had acted in good faith and had not violated their fiduciary duties to shareholders. However, in *Smith vs. Van Gorkom*, the Court held that the directors of Trans-Union Corporation breached their duty of care by approving a merger agreement without sufficient deliberation. This unexpected ruling had an immediate impact since it indicated that the Delaware Court would entertain the possibility of monetary damages against directors in situations where such damages were previously not thought to be applicable. The ruling contributed to a 34% increase in shareholder lawsuits in 1985 and an immediate escalation in liability insurance premiums for officers and directors (Wyatt 1988).

In response, in June of 1986, Delaware amended its corporate law to allow firms to enter into indemnification agreements with, and establish provisions to limit the personal liability of, their officers and directors. Numerous corporations rapidly took advantage of these provisions by reincorporating into Delaware. Although 32 other states had established similar statutes by 1988 (Pamepinto 1988), Delaware's quick action enabled it to capture 98% of the reincorporations, which were cited by management as being undertaken to reduce directors' liability, with more than half the reincorporating firms leaving California.

The 1987 proxy statement of Optical Coatings Laboratories is a good illustration of a proposal either to change its corporate charter in California or to reincorporate – to Delaware – for liability reasons, and documents the seriousness of the impact of liability insurance concerns on liability insurance premiums:

During 1986, the Company's annual premium for its directors' and officers' liability insurance was increased from \$17,500 to \$250,000 while the coverage was reduced from \$50,000,000 to \$5,000,000 in spite of the Company's impeccable record of never having had a claim. This is a result of the so-called directors' and officers' liability insurance crisis which has caused many corporations to lose coverage altogether and forced many directors to resign rather than risk financial ruin as a result of their good faith actions taken on behalf of their corporations.

This year at OCLI, we intend to do something about this problem. You will see included in the proxy materials a proposal to amend the Company's Articles of Incorporation, if California enacts the necessary legislation, to provide the Company's officers and directors with significantly greater protection from personal liability for their good faith actions on behalf of the Company. If California does not enact the necessary legislation by the date of the annual meeting, or any adjournment, a different proposal would provide for the Company to change its legal domicile to the State of Delaware, where the corporation law was recently amended to provide for such protection.

Although it was a Delaware Court decision that prompted the crisis in the director and officer liability insurance market, Delaware's quick action in remedying the situation by modifying its corporate laws reflects the general tendency for Delaware to be attentive to the changing needs of corporations. Romano (1985) contends that, because Delaware relies heavily upon corporate charter revenues, it has obligated itself to be an early mover in modifying its corporate laws to fit evolving business needs. It is clear that this tendency has proven beneficial in enhancing the efficiency of contracting for firms incorporating in Delaware.

In contrast to the reaction to the adoption of antitakeover measures, investors have responded positively to reincorporations that were undertaken to gain improved director liability protection. Observed abnormal stock returns averaging approximately +2.25% (again, at the 95% confidence level) are reported by Heron and Lewellen (1998). In a supplemental analysis, changes in the proportions of outside directors on the Boards of firms that reincorporated for director liability reasons were monitored for 2 years subsequent to the reincorporations, as a test of the claim that weak liability protection would make it more difficult for firms to attract outsiders to

their Boards. The finding was that firms that achieved director liability reduction via reincorporation did in fact increase their outside director proportions by statistically significant extents, whereas there was no such change for firms that reincorporated for other reasons.

19.4.3 Other Motives for Reincorporations

Reincorporations conducted solely to gain access to more flexible and predictable corporate laws, to save on taxes, to reconcile the firm's physical and legal domicile, and to facilitate acquisitions fall into the Contractual Efficiency category. Researchers have been unable to detect abnormal stock returns on the part of firms that have reincorporated for these reasons. The bulk of the reincorporations where managers cite the flexibility and predictability of the corporate laws of the destination state as motivation have been into Delaware. Romano (1985) argues that Delaware's responsive corporate code and its well-established set of court decisions have allowed the state to achieve a dominant position in the corporate chartering market. This argument would be consistent with the evidence that a substantial fraction of companies that reincorporate to Delaware do so just prior to an IPO of their stock. Indeed, Delaware has regularly chartered the lion's share of out-of-state corporations undergoing an IPO: 71% of firms that went public before 1991, 84% that went public between 1991 and 1995, and 87% of those that have gone public from 1996 (Bebchuk and Cohen 2002).

The language in the 1984 proxy statement of Computercraft provides an example of a typical proposal by management to reincorporate in order to have the firm take advantage of a more flexible corporate code:

The Board of Directors believes that the best interests of the Company and its shareholders will be served by changing its place of incorporation from the State of Texas to the State of Delaware. The Company was incorporated in the State of Texas in November 1977 because the laws of that state were deemed to be adequate for the conduct of its business. The Board of Directors believes that there is needed a greater flexibility in conducting the affairs of the Company since it became a publicly owned company in 1983.

The General Corporation Law of the State of Delaware affords a flexible and modern basis for a corporation action, and because a large number of corporations are incorporated in that state, there is a substantial body of case law, decided by a judiciary of corporate specialists, interpreting and applying the Delaware statutes. For the foregoing reasons, the Board of Directors believes that the activities of the Company can be carried on to better advantage if the Company is able to operate under the favorable corporate climate offered by the laws of the State of Delaware.

The majority of reincorporations which are done to realize tax savings or to reconcile the firm's legal domicile with its headquarters involve reincorporations

out of Delaware – not surprisingly, since Delaware is not only a very small state with few headquartered firms but also has annual chartering fees which are among the nation’s highest. The following excerpt from the 1989 proxy statement of the Longview Fibre Company illustrates the rationale for such a reincorporation:

Through the Change in Domicile, the Company intends to further its identification with the state in which the Company’s business originated, its principal business is conducted, and over 64% of its employees are located. Since the Company’s incorporation in the State of Delaware in 1926, the laws of the State of Washington have developed into a system of comprehensive and flexible corporate laws that are currently more responsive to the needs of businesses in the state.

After considering the advantages and disadvantages of the proposed Change in Domicile, the Board of Directors concluded that the benefits of moving to Washington outweighed the benefits and detriments of remaining in Delaware, including the continuing expense of Delaware’s annual franchise tax (the Company paid \$56,000 in franchise taxes in fiscal year 1988, whereas the “annual renewal fee” for all Washington corporations is \$50.00). In light of these facts, the Board of Directors believes it is in the best interests of the Company and its stockholders to change its domicile from Delaware to Washington.

Note in particular the issue raised about the annual franchise tax. Revenues from that source currently account for approximately \$400 million of Delaware’s state budget (Bebchuk and Cohen 2002).

19.5 Summary and Conclusions

Distinctive among major industrialized countries, incorporation in the U.S. is a state rather than a federal process. Hence, there are a wide variety of legal domiciles that an American firm can choose from, and the corporation laws of those domiciles vary widely as well – in areas such as the ability of shareholders to hold a firm’s managers accountable for their job performance, the personal liability protection afforded to corporate officers and directors, and the extent to which management can resist attempts by outsiders to take over the firm. The resulting array of choices of chartering jurisdictions has been characterized by two competing views: (1) the diversity is desirable because it enables a firm to select a legal domicile whose laws provide the most suitable and most efficient set of contracting opportunities for the firm’s particular circumstances; (2) the diversity is undesirable because it encourages states to compete for incorporations – and reincorporations – by passing laws that appeal to a firm’s managers by insulating them from shareholder pressures and legal actions, and making it difficult for the firm to be taken over without management’s concurrence. Thus, the choice of legal domicile can become an important element in the governance of the firm, and a *change* of domicile can be a significant event for the firm.

As for many other aspects of corporate decision-making, a natural test as to which of the two characterizations are correct is to observe what happens to the stock prices of companies who reincorporate, on and around the time they do so. The available evidence indicates that reincorporations which result in the firm gaining additional takeover defenses have negative impacts on its stock price – apparently, because investors believe that a takeover and its associated premium price for the firm’s shares will thereby become less likely. Conversely, reincorporations that occasion an increase in the personal liability protection of officers and directors have positive stock price effects. The inference is that such protection makes it easier for the firm to attract qualified directors who can then help management improve the firm’s financial performance. These effects are accentuated when the reincorporation is accompanied by a clear statement from management to the firm’s shareholders about the reasons for the proposed change. There is, therefore, some support for both views of the opportunity for firms to “shop” for a legal domicile, depending on the associated objective. Other motives for reincorporation seem to have little if any impact on a firm’s stock price, presumably because they are not regarded by investors as material influences on the firm’s performance.

References

- Baysinger, B.D., and H.N. Butler. 1985. The role of corporate law in the theory of the firm. *Journal of Law and Economics* 28: 179–191.
- Bebchuk, L. 1992. Federalism and the corporation: The desirable limits on state competition in corporate law. *Harvard Law Review* 105: 1435–1510.
- Bebchuk, L.A., and A. Cohen. 2002. *Firms’ decisions where to incorporate. Discussion paper no. 351*. Cambridge, MA: Harvard Law School, Olin Center for Law, Economics, and Business, Feb 2002. www.law.harvard.edu/programs/olin_center
- . 2003. Firms’ decisions where to incorporate. *Journal of Law and Economics* 46: 383–425.
- Bebchuk, L.A., and A. Ferrell. 1999. Federalism and corporate law: The race to protect managers from takeovers. *Columbia Law Review* 99: 1168–1199.
- Cary, W. 1974. Federalism and corporate law: Reflections upon Delaware. *Yale Law Journal* 83: 663–707.
- Daines, R., and M. Klausner. 2001. Do IPO charters maximize firm value? An empirical examination of antitakeover defenses in IPOs. *Journal of Law, Economics, and Organization* 17: 83–120.
- Dodd, P., and R. Leftwich. 1980. The market for corporate charters: “Unhealthy competition” versus federal regulation. *Journal of Business* 53: 259–282.
- Easterbrook, F.H., and D.R. Fischel. 1983. Voting in corporate law. *Journal of Law and Economics* 26: 395–427.
- Field, L., and J. Karpoff. 2002. Takeover defenses of IPO firms. *Journal of Finance* 57: 1857–1889.
- Heron, R., and W. Lewellen. 1998. An empirical analysis of the reincorporation decision. *Journal of Financial and Quantitative Analysis* 33: 549–568.

-
- Jensen, M., and W. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics* 3: 305–360.
- Pamepinto, S. 1988. *Reincorporation: 1988 background report*. Investor Responsibility Research Center Report, Washington, DC.
- Romano, R. 1985. Law as a product: Some pieces of the reincorporation puzzle. *Journal of Law, Economics, and Organization* 12: 225–283.
- Wyatt Company. 1988. *Directors and officers liability survey*. Chicago: Wyatt.



Mean Variance Portfolio Allocation

20

Cheng Hsiao and Shin-Huei Wang

Abstract

The basic rules of balancing the expected return on an investment against its contribution to portfolio risk are surveyed. The related concept of Capital Asset Pricing Model asserting that the expected return of an asset must be linearly related to the covariance of its return with the return of the market portfolio if the market is efficient and its statistical tests in terms of Arbitrading Price Theory are also surveyed. The intertemporal generalization and issues of estimation errors and portfolio choice are discussed as well.

Keywords

Arbitrage pricing theory · Capital asset pricing model · Covariance · Errors of estimation · Mean–variance efficiency · Minimum variance portfolio · Sharpe ratio · Value at risk · Volatility · Zero-beta portfolio

20.1 Introduction

Stock prices are volatile. The more volatile a stock, the more uncertain its future value. Investment success depends on being prepared for and being willing to take risk. The insights provided by modern portfolio theory arise from the interplay between the mathematics of return and risk. The central theme of modern portfolio theory is: “In constructing their portfolios investors need to look at the expected

C. Hsiao (✉)

University of Southern California, Los Angeles, CA, USA

e-mail: chsiao@usc.edu

S.-H. Wang

University of Southern California, Los Angeles, CA, USA

The Center for Operations Research and Econometrics (CORE), Louvain-la-Neuve, Belgium

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,

https://doi.org/10.1007/978-3-030-91231-4_20

743

return of each investment in relation to the impact that it has on the risk of the overall portfolio” (Litterman et al. 2003).

To balance the expected return of an investment against its contribution to portfolio risk, an investment’s contribution to portfolio risk is not just the risk of the investment itself, but rather the degree to which the value of that investment moves up and down with the values of the other investments in the portfolio. This degree to which these returns move together is measured by the statistical quantity called “covariance,” which is itself a function of their correlation along with their volatilities when volatility of a stock is measured by its standard deviation (square root of variance). However, covariances are not observed directly; they are inferred from statistics that are notoriously unstable.

In Sect. 20.2, we summarize the Markowitz (1952, 1959) mean–variance allocations rule under the assumption that the correlations and volatilities of investment returns are known. Section 20.3 describes the relationship between the mean variance efficiency and asset pricing models. Section 20.4 discusses issues of estimation in portfolio selection.

20.2 Mean–Variance Portfolio Selection

The basic portfolio theory is normative. It considers efficient techniques for selecting portfolios based on predicted performance of individual securities. Marschak (1938) was the first to express preference in terms of indifference curves in a mean–variance space. Von Neumann and Morgenstern (1947) provided an axiomatic framework to study the theory of choice under uncertainty. Based on these developments, Markowitz (1952, 1959) developed a mean–variance approach of asset allocation.

Suppose there are N securities indexed by i , $i = 1, \dots, N$. Let $R' = (R_1, \dots, R_N)$ denote the return of these N securities. Let $\mu = ER$ and Σ be the mean and the nonsingular covariance matrix of R . A portfolio is described by an allocation vector $X' = (x_1, \dots, x_N)$ of quantity x_i for the i th security. In the mean–variance approach, an investor selects the composition of the portfolio to maximize her expected return while minimizing the risk (i.e. the variance) subject to budget constraint. Since these objectives are contradictory, the investor compromises and selects the portfolio that minimizes the risk subject to a given expected return, say d . A portfolio X is said to be the minimum-variance portfolio of all portfolios with mean (or expected) return d if its portfolio weight vector is the solution to the following constrained minimization:

$$\min_x X' \sum X \quad (1)$$

subject to

$$X' \mu = d, \quad (2)$$

and

$$X' i = 1, \quad (3)$$

where i is an $N \times 1$ vector of ones. Solving the Lagrangian

$$L = X' \sum X + \lambda_1 [d - X' \mu] + \lambda_2 (1 - X' i), \quad (4)$$

yields the optimal portfolio

$$X_p = \frac{1}{D} \left\{ \left[\left(\mu' \sum^{-1} \mu \right) \sum^{-1} i - \left(i' \sum^{-1} \mu \right) \sum^{-1} \mu \right] + d \left[\left(i' \sum^{-1} i \right) \sum^{-1} \mu - \left(i' \sum^{-1} \mu \right) \sum^{-1} i \right] \right\}, \quad (5)$$

where

$$D = \left(\mu' \sum^{-1} \mu \right) \left(i' \sum^{-1} i \right) - \left(i' \sum^{-1} \mu \right)^2 \quad (6)$$

From Eq. (5), we have:

Proposition 1

Any two distinct minimum-variance portfolios can generate the minimum variance frontier.

Proposition 2

Let portfolio o as the portfolio with the smallest possible variance for any mean return, then the global minimum variance portfolio has

$$X_0 = \frac{1}{\left(i' \sum^{-1} i \right)} \sum^{-1} i, \mu_0 = ER_0 = EX'_0 R = \frac{i' \sum^{-1} \mu}{i' \sum^{-1} i}, = \frac{1}{i' \sum^{-1} i} \quad (7)$$

Proposition 3

The covariance of the return of the global minimum-variance portfolio o with any portfolio p is

$$\text{Cov}(R_o, R_p) = \frac{1}{i' \sum^{-1} i}. \quad (8)$$

that is, the correlation of them is positive, $\text{corr}(R_o, R_p) > 0$ for any portfolio p.

Proposition 4

If the covariance of the returns of two portfolios p and q equal to 0, $\text{Cov}(R_p, R_q) = 0$, then portfolios p and q are called orthogonal portfolios and the portfolio q is the unique portfolio which is orthogonal to p.

Proposition 5

All portfolios on positively slope part of mean-variance frontier are positively correlated.

When a risk-free asset with return R_f is present, the expected return of investing in the $N + 1$ assets will be

$$\mu_0 = (1 - \alpha)R_f + \alpha X' \mu, \quad (9)$$

where $0 < \alpha < 1$ is the proportion of investment in the risky assets. The minimum-variance portfolio with the expected return of investing in both N risky assets and the risk-free asset equal to d , $\mu_\alpha = d$, is the solution of

$$\alpha^2 X' \Sigma X \quad (10)$$

subject to Eqs. (3) and (9) equal to d ,

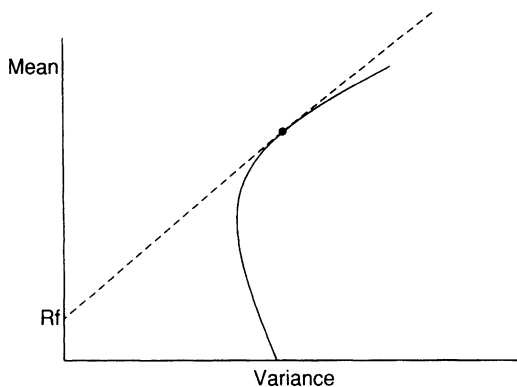
$$X_p^* = \frac{1}{i' \Sigma^{-1} (\mu - R_f i)} \Sigma^{-1} (\mu - R_f i), \quad (11)$$

$$\alpha = \frac{i' \Sigma^{-1} (\mu - R_f i)}{(\mu - R_f i) \Sigma^{-1} (\mu - R_f i)} (d - R_f). \quad (12)$$

Thus, when there is a risk-free asset, all minimum-variance portfolios are a combination of a given risky asset portfolio with weights proportional to X_p^* and the risk-free asset. This portfolio of risky assets is called the tangency portfolio because X_p^* is independent of the level of expected return. If we draw the set of minimum-variance portfolios in the absence of a risk-free asset in a two-dimensional mean-standard deviation space like the curve GH in Fig. 1, all efficient portfolios lie along the line from the risk-free asset through portfolio X_p^* .

Sharpe (1964) proposes a measure of efficiency of a portfolio in terms of the excess return per unit risk. For any asset or portfolio with an expected return μ_α and standard deviation σ_α , the *Sharpe ratio* is defined as

Fig. 1 Mean-variance frontier with risk-free asset



$$Sr_a = \frac{\mu_a - R_f}{\sigma_a} \quad (13)$$

The tangency portfolio X_p^* is the portfolio with the maximum *Sharpe ratio* of all risky portfolios. Therefore, testing the mean–variance efficiency of a given portfolio is equivalent to testing if the *Sharpe ratio* of that portfolio is maximum of the set of *Sharpe ratios* of all possible portfolios.

20.3 Mean–Variance Efficiency and Asset Pricing Models

20.3.1 Capital Asset Pricing Models

The Markowitz mean–variance optimization framework is from the perspective of individual investor conditional on given expected excess returns and measure of risk of securities under consideration. The Capital Asset Pricing Model (CAPM) developed by Sharpe (1964) and Lintner (1965) asks what values of these mean returns will be required to clear the demand and supply if markets are efficient, all investors have identical information, and investors maximize the expected return and minimize volatility.

An investor maximizing the expected return and minimizing risk will choose portfolio weights for which the ratio of the marginal contribution to portfolio expected return to the marginal contribution to risk will be equal. In equilibrium, expected excess returns are assumed to be the same across investors. Therefore, suppose there exists a risk-free rate of interest, R_f , and let $Z_i = R_i - R_f$ be the excess return of the i th asset over the risk-free rate (R_f) then the expected excess return for the i th asset in equilibrium is equal to

$$E[Z_i] = \beta_{im}E[Z_m], \quad (14)$$

and

$$\beta_{im} = \frac{\text{Cov}(Z_i, Z_m)}{\text{Var}(Z_m)}, \quad (15)$$

where Z_m is the excess return on the market portfolio of assets, $Z_m = R_m - R_f$, with R_m being the return on the market portfolio.

In the absence of a risk-free asset, Black (1971) derived a more general version of the CAPM. In the Black version, the expected return of asset i in equilibrium is equal to

$$E[R_i] = E[R_{om}] + \beta_{im}\{E[R_m] - E[R_{om}]\}, \quad (16)$$

and

$$\beta_{im} = \frac{\text{Cov}(R_i, R_m)}{\text{Var}(R_m)}, \quad (17)$$

where R_{om} is the return on the *zero-beta portfolio* associated with m . The *zero-beta portfolio* is described as the portfolio that has the minimum variance among all portfolios that are uncorrelated with m .

Closely related to the concept of trade-off between risk and expected return is the quantification of this trade-off. The CAPM or *zero-beta CAPM* provides a framework to quantify this relationship. The CAPM implies that the expected return of an asset must be linearly related to the covariance of its return with the return of the market portfolio and the market portfolio of risky assets is a mean–variance efficient portfolio. Therefore, studies of market efficiency have been cast in the form of testing the Sharpe–Lintner CAPM and the *zero-beta CAPM*.

Under the assumption that returns are independently, identically (IID) multivariate normally distributed, empirical tests of the Sharpe–Lintner CAPM have focused on the implication of Eq. (14) that the regression of excess return of the i th asset at time t , $Z_{it} = R_{it} - R_{ft}$ on the market excess return at time t , $Z_{mt} = R_{mt} - R_{ft}$, has intercept equal to zero. In other words for the regression model

$$Z_{it} = c_{im} + \beta_{im}Z_{mt} + \varepsilon_{it}, \quad \begin{matrix} i = 1, \dots, N, \\ t = 1, \dots, T. \end{matrix} \quad (18)$$

The null hypothesis of market portfolio being mean–variance efficient is:

$$H_0 : c_{1m} = c_{2m} = \dots = c_{Nm} = 0. \quad (19)$$

Empirical tests of Black (1971) version of the CAPM model note that Eq. (15) can be rewritten as

$$\begin{aligned} E[R_i] &= \alpha_{im} + \beta_{im}E[R_m], \\ \alpha_{im} &= E[R_{om}](1 - \beta_{im}) \forall i. \end{aligned} \quad (20)$$

That is, the Black model restricts the asset-specific intercept of the (inflation adjusted) real-return market model to be equal to the expected *zero-beta* portfolio return times one minus the asset's beta. Therefore, under the assumption that the real-return of N assets at time t , $R_t = (R_{1t}, \dots, R_{Nt})'$ is IID (independently identically distributed) multivariate normal, the implication of the Black model is that the intercepts of the regression models

$$R_{it} = \alpha_{im} + \beta_{im}R_{mt} + \varepsilon_{it}, \quad \begin{matrix} i = 1, \dots, N, \\ t = 1, \dots, T. \end{matrix} \quad (21)$$

are equal to

$$\alpha_{im} = (1 - \beta_{im})\gamma, \quad (22)$$

where the constant γ denotes the expected return of *zero-beta portfolio*.

20.3.2 Arbitrage Pricing Theory

Although CAPM model has been the major frame-work for analyzing the cross-sectional variation in expected asset returns for many years, Gibbons (1982) could not find empirical support for the substantive content of the CAPM using stock returns from 1926 to 1975. His study was criticized by a number of authors from both the statistical methodological point of view and the empirical difficulty of estimating the unknown zero beta return (e.g. Britten-Jones 1999; Campbell et al. 1997; Gibbons et al. 1989; Shanken 1985; Stambaugh 1982; Zhou 1991). Ross (1977) notes that no correct and unambiguous test can be constructed because of our inability to observe the exact composition of the true market portfolio. Using arbitrage arguments, Ross (1977) proposes the Arbitrage Pricing Theory (APT) as a testable alternative. The advantage of the APT is that it allows for multiple risk factors. It also does not require the identification of the market portfolio.

Under the competitive market the APT assumes that the expected returns are functions of an unknown number of unspecified factors, say $K(K < N)$:

$$R_{it} = E_i + \beta_{i1}\delta_{1t} + \dots + \beta_{iK}\delta_{Kt} + \varepsilon_{it}, \quad \begin{matrix} i = 1, \dots, N, \\ t = 1, \dots, T, \end{matrix} \quad (23)$$

where R_{it} is the return on asset i at time t , E_i is its expected return, β_{ik} are the factor loadings, δ_{kt} are independently distributed zero mean common factors and the ε_{it} are zero mean asset-specific disturbances, assumed to be uncorrelated with the δ_{kt} .

As an approximation for expected return, the APT is impossible to reject because the number of factors, K , is unknown. One can always introduce additional factors to satisfy Eq. (23). Under the additional assumption that market portfolios are well diversified and that factors are pervasive, Connor (1984) shows that it is possible to have exact factor pricing. Dybvig (1985) and Grinblatt and Titman (1987), relying on the concept of “local mean–variance efficiency,” show that given a reasonable specification of the parameters of an economy, theoretical deviations from exact factor pricing are likely to be negligible. Thus, the factor portfolios estimated by the maximum likelihood factor analysis are locally efficient if and only if the APT holds (Roll and Ross 1980; Shanken 1982; Dybvig and Ross 1985).

20.3.3 Intertemporal Capital Asset Pricing Model (ICAPM)

The multifactor pricing models can alternatively be derived from an intertemporal equilibrium argument. The CAPM models are static models. They treat asset prices as being determined by the portfolio choices of investors who have preferences defined over wealth after one period. Implicitly, these models assume that investors consume all their wealth after one period. In the real world, investors consider many periods in making their portfolio decisions. Under the assumption that consumers maximize the expectation of a time-separable utility function and use financial assets to transfer wealth between different periods and states of the world and relying on the

argument that consumers' demand is matched by the exogenous supply, Merton (1973) shows that the efficient portfolio is a combination of one of mean–variance efficient portfolio with a hedging portfolio that reflects uncertainty about future consumption investment state. Therefore, in Merton Intertemporal Capital Asset Pricing Model (ICAPM) it usually lets market portfolio serve as one factor and state variables serve as additional factors.

The CAPM implies that investors hold a mean–variance portfolio that is a tangency point between the straight line going through the risk-free return to the minimum-variance portfolios without risk-free asset in the mean-standard deviation space. Fama (1996) shows that similar results hold in multifactor efficient portfolios of ICAPM.

20.4 Estimation Errors and Portfolio Choice

The use of mean variance analysis in portfolio selection requires the knowledge of means, variances, and covariances of returns of all securities under consideration. However, they are unknown. Treating their estimates as if they were true parameters can lead to suboptimal portfolio choices (e.g. Frankfurter et al. 1971; Klein and Bawa 1976; Jorion 1986) have conducted experiments to show that because of the sampling error, portfolios selected according to the Markowitz criterion are no more efficient than an equally weighted portfolio. Chopra (1991), Michaud (1989), and others have also shown that mean–variance optimization tends to magnify the errors associated with the estimates.

Chopra and Ziemba (1993) have examined the relative impact of estimation errors in means, variances, and covariances on the portfolio choice by a measure of percentage cash equivalent loss (CEL). For a typical portfolio allocation of large U.S. pension funds, the effects of CEL for errors in means are about 11 times as that of errors in variances and over 20 times as that of errors in covariances. The sensitivity of mean-variance efficient portfolios to changes in the means of individual assets was also investigated by Best and Grauer (1991) using a quadratic programming approach.

The main argument of these studies appears to be that in constructing an optimal portfolio, good estimates of expected returns are more important than good estimates of risk (covariance matrix). However, this contention is challenged by De Santis et al. (2003). They construct an example showing the estimates of Value at Risk (VaR), identified as the amount of capital that would be expected to be lost at least once in 100 months, using a \$100 million portfolio invested in 18 developed equity markets to the sensitivity of different estimates of covariance matrix. Two different estimates of the covariance matrix of 18 developed equity markets are used – estimates using equally weighted 10 years of data and estimates giving more weights to more recent observations. They show that changes in estimated VaR can be between 7% and 21%.

The main features of financial data that should be taken into account in estimation as summarized by De Santis et al. (2003) are:

1. Volatilities and correlations vary over time.
2. Given the time-varying nature of second moments, it is preferable to use data sampled at high frequency over a given period of time, rather than data sampled at low frequency over a longer period of time.
3. When working with data at relatively high frequencies, such as daily data, it is important to take into account the potential for autocorrelations in returns.
4. Daily returns appear to be generated by a distribution with heavier tails than the normal distribution. A mixture of normal distributions appear to approximate the data-generating process well.
5. Bayesian statistical method can be a viable alternative to classical sampling approach in estimation (e.g. Jorion 1986).

References

- Best, M.J., and R.R. Grauer. 1991. On the sensitivity of mean variance efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies* 4 (2): 315–342.
- Black, F. 1971. Capital market equilibrium with restricted borrowing. *Journal of Business* 45: 444–454.
- Britten-Jones, M. 1999. The sampling error in estimates of mean efficient portfolio weight. *Journal of Finance* 54 (2): 655–671.
- Campbell, J.Y., A.W. Lo, and A.C. MacKinley. 1997. *The econometrics of financial markets*. Princeton: Princeton University Press.
- Chopra, Vijay K. 1991. Mean variance revisited: Near optimal portfolios and sensitivity to input variations. *Russell Research Commentary*.
- Chopra, V.K., and W.T. Ziemba. 1993. The effect of errors in mean, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management* 19: 6–11.
- Connor, G. 1984. A unified beta pricing theory. *Journal of Economic Theory* 34: 13–31.
- De Santis, G., B. Litterman, A. Vesval, and K. Winkelmann. 2003. Covariance matrix estimation. In *Modern investment management*, ed. B. Litterman and the Quantitative Resources Group, Goldman Sachs Asset Management. Hoboken: Wiley.
- Dybvig, P.H. 1985. An explicit bound on individual assets' deviation from APT pricing in a finite economy. *Journal of Financial Economics* 12: 483–496.
- Dybvig, P.H., and S.A. Ross. 1985. Yes, the APT is testable. *Journal of Finance* 40: 1173–1183.
- Fama, E.F. 1996. Multifactor portfolio efficiency and multifactor asset pricing. *Journal of Financial and Quantitative Analysis* 31 (4): 441–465.
- Frankfurter, G.M., H.E. Phillips, and J.P. Seagle. 1971. Portfolio selection: The effects of uncertain means, variances, and covariances. *Journal of Financial and Quantitative Analysis* 6 (5): 1251–1262.
- Gibbons, M.R. 1982. Multivariate tests of financial models: A new approach. *Journal of Financial Economics* 10: 3–27.
- Gibbons, M.R., S.A. Ross, and J. Shanken. 1989. A test of efficiency of a given portfolio. *Econometrica* 57: 1121–4152.
- Grinblatt, M., and S. Titman. 1987. The relation between mean-variance efficiency and arbitrage pricing. *Journal of Business* 60 (1): 97–112.
- Jorion, P. 1986. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21 (3): 279–292.
- Klein, R.W., and V.S. Bawa. 1976. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics* 3: 215–231.

- Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47 (1): 13–37.
- Litterman, B., and the Quantitative Resources Group, Goldman Sachs Asset Management. 2003. *Modern investment management*. Hoboken: Wiley.
- Markowitz, H.M. 1952. Portfolio selection. *The Journal of Finance* 7: 77–91.
- . 1959. *Portfolio selection: Efficient diversification of investments*. New York: Wiley.
- Marschak, J. 1938. Money and the theory of assets. *Econometrica* 6: 311–325.
- Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867–887.
- Michaud, R.O. 1989. The Markowitz optimization enigma: Is “optimized” optimal? *Financial Analysts Journal* 45: 31–42.
- Roll, R., and S.A. Ross. 1980. An empirical investigation of the arbitrage pricing theory. *Journal of Finance* 33: 1073–1103.
- Ross, R. 1977. A critique of the asset pricing theory’s tests, part 1: One past and potential testability of the theory. *Journal of Financial Economics* 4: 129–176.
- Shanken, J. 1982. The arbitrage pricing theory: Is it testable? *Journal of Finance* 37 (5): 1129–1140.
- . 1985. Multivariate tests of the zero-beta. *Journal of Financial Economics* 14: 327–348.
- Sharpe, W.F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- Stambaugh, R.F. 1982. On the exclusion of assets from tests of the two parameter model. *Journal of Financial Economics* 10: 235–268.
- Von Neumann, J., and O. Morgenstern. 1947. *Theory of games and economic behavior*. 2nd ed. Princeton: Princeton University Press.
- Zhou, G. 1991. Small samples tests of portfolio efficiency. *Journal of Financial Economics* 30: 165–191.



Chang-Tseh Hsieh

Abstract

The proliferation of the Internet has led to the rapid growth of online brokerage. As the Internet now allows individual investors access to information previously available only to institutional investors, individual investors are profiting in the financial markets through online trading schemes. Rock-bottom fees charged by the online brokers and the changing attitude toward risk of the Internet-literate generation prompt the practitioners to question the validity of the traditional valuation models and statistics-based portfolio formulation strategies. These tactics also induce more dramatic changes in the financial markets. Online trading, however, does involve a high degree of risk, and can cause a profitable portfolio to sour in a matter of minutes. This chapter addresses the major challenges of trading stocks on the Internet, and recommends a decision support system for online traders to minimize the potential of risks.

Keywords

Day trading center · Decision support systems · Financial market · Institutional investors · Internet · Online trading · Portfolio formulation strategies · Risks management · Stock investment · Valuation models · Web-based brokers

21.1 Introduction

In the past decade, one of the most phenomenal changes in investment markets is the burgeoning number of online brokers, and its subset, the so-called day-trading centers. Instead of doing business using the old style face-to-face approach, or over the phone with stockbroker, investors have been using the Web to explore a

C.-T. Hsieh (✉)

University of Southern Mississippi, Hattiesburg, MS, USA

e-mail: chang-tseh.hsieh@usm.edu

wealth of free information and have been making investment decisions with a new fleet of Internet-based brokers. Concepts of online trading have been around for quite some time. Before the proliferation of the Internet, however, online trading was primarily used as a vehicle for trading by institutional investors. With the help of the Internet, individual investors are now able to access the stock markets in ways similar to those of the major players, the institutional investors (Barnett 1999b; Smith 1999a). The direct use of the Internet to trade stocks also raises doubt among investors about the validity of the traditional stock valuation models as well as portfolio formulation strategies.

Inspired by the successful story of E* Trade, the pioneering Internet-based broker, many Web-based brokers have joined the throng that has forced traditional full-service firms to respond with bigger changes. Although, by the end of 1998, online brokers still controlled only \$400 billion of assets in customer accounts as compared with \$3200 billion managed by full-service brokers, transactions done through online traders now represent more than 15% of all equity trades, a twofold increase in just 2 years. And the online brokerage industry has doubled customer assets to more than \$420 billion, and doubled accounts managed to 7.3 million by early 1999.

The Internet has revolutionized the way in which consumers perform research and participate in the buying and selling of securities. As of January 2003, there are an estimated 33 million U.S. consumer online trading accounts that control roughly \$1.6 trillion in customer assets (Intel International Group Ltd 2002). The convenience of online trading has introduced millions of new consumers to the possibilities of online money management. At the same time, the Internet and wireless devices have transformed the way in which capital markets operate and have made it possible for individual investors to have direct access to a variety of different markets, and to tools that were at one time reserved only for the investment professional.

21.2 The Issues

The proliferation of online trading sites has created major changes in the ways stocks are traded. Traditionally, an investor who wants to purchase a stock has to go through a broker. The broker will send a buy order to a specialist on the exchange floor, if the stock is listed on the NYSE. The specialist then looks for sellers on the trading floor or in his electronic order book. If the specialist finds enough sellers to match his offer price, the specialist completes the transaction. Otherwise, the specialist may purchase at a higher price, with customer permission, or sell the stock to the customer out of his own inventory.

If the stock is listed on the NASDAQ, the broker consults a trading screen that lists offers from the market makers for the said stock. The broker then picks up the market maker with the best price to complete the transaction. On the other hand, for buying stock online, the broker such as E* Trade simply collects order information, and completes the transaction through the electronic communication network (ECN).

For traditional brokerage services, the broker usually charges hefty fees. For example, Morgan Stanley Dean Witter charges \$40 per trade for customers with at least \$100,000 in their accounts and if they make at least 56 trades per year. Merrill Lynch charges \$56 per trade for a \$100,000 account, with 27 trades per year. Typically, fees for a single trade at the full-service brokerage can be anywhere from \$100 to \$1000 depending upon the services involved. Charles Schwab, however, charges \$29.95 per trade up to 1000 shares, and the champion of the online trader, E* Trade, charges merely \$14.95 per market-order trade up to 5000 shares. Alternatively, the investor can choose unlimited number of trades and access to exclusive research and advice for a yearly fee (Thornton 2000).

The reduced cost offered by online trading has encouraged investors to increase the frequency of trading. Since the fee paid to complete a transaction through traditional brokerage is enough to cover fees of many trades charged online, the investor can afford to ride the market wave to try and realize a windfall caused by the price fluctuation on a daily basis. Perhaps, this helps explain why in two short years, Island, Instinet, and seven other ECNs, now control a whopping 21.6% of NASDAQ shares and nearly a third of the trades and are seeking to expand their operations to include NYSE company shares (Vogelstein 1999a; Reardon 2000).

Although investors of all sizes could use online brokers, the most noteworthy change in financial markets is the increasing number of individual investors. These are the new breed of investors armed with the knowledge of information technology and a very different attitude toward risk in the investment market place (Pethokoukis 1999a, b). Their changing attitudes have contributed to several major changes in stock market strategies (Becker 1998; Barnett 1999a; Gimein 1999; Pethokoukis 1999a, b; Vogelstein 1999b; Sharma 2000).

1. *Webstock frenzies.* Although day traders represent a small percentage of all active traders on a daily market, the industry makes up about 15% of NASDAQ's daily volume (Smith 1999b). The aggressive trading behavior of day traders, fueled by margin loans supplied by day-trading centers, is one of the driving forces behind the runaway price of many Internet-related stocks. Since the beginning of 1999, for example, Yahoo stock rose \$40 in 1 day. eBay shares fell \$30. [Broadcast.com](#) gained \$60 a share, and then lost \$75 2 days later. Webstocks as a whole gained 55% in the first days of trading in 1999, then a free-fall started in the early summer. Since April 1999, American Online stock price has dropped almost 67%. And the Goldman Sachs Internet Index currently stands nearly 43% below its all-time high in April. By Spring 2000, many of the tech stocks have recorded more than 80% of share price corrections. Some of these corrections actually happened in just a matter of few days (Cooper 2001). These stocks have taught the online/day traders the real meaning of "volatility" (McLean 1999).
2. *Changing goals of investment.* The easy money mentality has led to new goals for formulating investment portfolios. Traditional portfolio models have been based on a mean-variance modeling structure, and for years numerous variations of such models have filled the academic journals. Today, however, investment professionals have been forced to abandon the investment strategies developed by

academics, focusing instead on strategies that achieve instant profits. As Net stocks became the horsepower to help pump the DJ index near to the 11,000 mark by early 2000, investors have renounced traditional buy-and-hold strategies and have switched to holding stocks for minutes at a time. In addition, the changing investment goals are partially caused by the change in the valuation system.

3. *Different valuation models.* Many of today's hot stocks are not worth anywhere near where they trade. For example, Netstock Amazon.com, one of the hottest, sold just \$610 million in books and CDs in 1999 and is yet to make its first penny. However, its \$20 billion market value makes it worth \$5 billion more than Sears. In fact, with the exception of Yahoo, all Webstocks have infinite P/E ratios. This anomaly prompts practitioners to question traditional models of valuing the stock, and forecasters everywhere concede that old models are suspect (Weber 1999).

Online trading also engenders some changes in the traditional investing scenario. First, the wide variation in investor knowledge of the stock market and of trading is crucial in the online setting. The costs to investors of bad judgment are likely to be borne by new entrants to the world of individual investing; these investors are pleased with the simplicity of the interactive user-friendly formats of e-brokerages, but are seldom proficient in the mechanisms and arrangements beyond the interface. Experienced investors can better identify the benefits and costs of choosing specific e-brokerages.

Second, the frequency of online investor trading deserves special attention. Many market analysts suggest that the growing U.S. economy and the low commissions charged by e-brokerages influence investors to trade more often. For example, an average Merrill Lynch (full-service broker) customer makes four to five trades per year while the core investors in an e-brokerage such as E* Trade make an average of 5.4 trades per quarter. Frequent trading is generally contrary to the recommendations of financial theory. Ultimately, it is possible for an e-brokerage to allow investors to trade frequently at very low or even zero costs per trade while earning large profits on the fraction of the increasingly large bid-ask spread that is pushed back by the market maker. At the same time, the investor may be unaware of the indirect costs incurred with each trade.

Third, the evolution of electronic trading may increase market fragmentation in the short run. E-brokerages may increasingly channel trades away from exchanges and toward market makers to compensate for lost revenue resulting from low direct commissions. Market fragmentation may have a negative impact on prices, increasing the bid-ask spread and potential for arbitrage opportunities (e.g. buy low in one market and sell high in another market within a short period of time). This is contrary to the belief that electronic markets may force centralization and increase liquidity (i.e. the ability to buy and sell securities quickly).

Clearly, people's attitudes toward risk have been changing constantly. Many behavioral factors that have not been successfully incorporated into traditional quantitative models have now become decisive factors in valuing investments. Several new models have thus been developed in an effort to better explain why all of a sudden investors do not see the stock market as the dangerous place they once did.

21.3 Some New Portfolio Structure Models

Among the new models, which overturn statistical relationships that have held true for decades, the major ones are (Glassman and Hassett 1998):

1. *Fed model*. Edward Yardeni, an economist at the Deutsche Bank, developed this model. The model relates earnings yield on stocks to interest rates. When the earnings yield is equal to the current yield on a 10-year U.S. Treasury bond, stocks are at fair value. If the earnings yield is above the interest rate, stocks are a buy; if below, stocks are overvalued.
For example, over the next 12 months, the consensus earnings forecast of industry analysts for the S&P 500 is \$52.78 per share. This is a 19.1% increase over the latest available four-quarter trailing sum of earnings. The fair value of the S&P 500 Index was 1011.11, derived as the 12-month forward earnings divided by the 10-year Treasury bond yield, assuming at 5.22%. If the S&P 500 closed at 1318.31, then the market would be 30.4% overvalued. Individual investors can enter their projected bond yield and estimated growth in corporate earnings to check the valuation of the stocks at Yardeni's Web site.
2. *Campbell–Shiller model*. The valuation model developed by John Y. Campbell of Harvard University and Robert J. Shiller of Yale University looks at price earnings ratios over time to determine a long-term market average (Campbell 1987, 1996; Campbell and Shiller 1991). When the current P/E exceeds that average, the market is overvalued. For example, the long-term average of P/E is 15. Therefore, at its current ratio of over 33, the stock price is overvalued.
3. *Cornell model*. This model discounts future cash flows and compares that to the current market level. The discount factor is a combination of the risk-free interest rate and a risk premium to compensate for the greater volatility of stocks. When the value of the discounted cash flows is above the current price, the market is cheap. Otherwise, it is overvalued.
4. *Glassman–Hassett model*. Similar to the Cornell model with one major exception, Glassman and Hassett argue that the risk premium, historically at 7%, is heading toward 0%. This means the discount factor that applies to stocks drops sharply, thus raising the fair value of the market.

The Internet has drastically changed the way investors make investment decisions. Technology empowers individual investors through many inventions and innovated services. Much information traditionally available only to institutional investors is now accessible to individual investors through the World Wide Web. For example, the Thomson Investors site allows individual traders to view the institutional pre-trade activity and get a bird's eye view of the activity on the NYSE floor. One very useful source of information is from StarMine. Investors can use this Web site to identify experts worth listening to, then use Multex to get the full detail of the relevant information (Mullaney 2001). At Bestcalls.com, visitors can examine conference call information, and in the near future, individuals will be able to see corporate officers

deliver the bullet points of their business models to institutions at www.eoverview.com, Net Roadshow's Web site.

These services may not bring individual investors up to par with institutional investors. However, they are now able to make investment decisions based on information with similar quality and currency as the big investors. The saved costs of trading through online brokers might provide individual investors an edge over their big counterparts. Since the individual investors' activity usually involves only small volumes of a given stock, their decisions will not likely cause a great fluctuation in the price. This will enable them to ride the market movement smoothly. Nevertheless, online traders must be aware that not all online brokers are competent. It is very important to sign up services with reputed brokers, who are backed with solid Internet infrastructure to minimize the frustration with those brokers (Gogoi 2000).

Trading online, however, involves an unusually high degree of risks. Since most online traders are looking for profits in a relatively short period of time, their investing targets are primarily in tech-concentrated NASDAQ markets where volatility is the rule (McNamee 2000; Opiela 2000). Yet, many online investors forget that online or off, disciplines for managing portfolio to minimize risk are still indispensable (Campbell 1996; Brockman and Chung 2000; Farrell 2000).

21.4 Conclusion

Online trading provides convenience, encourages increased investor participation, and leads to lower upfront costs. In the long run, these will likely reflect increased market efficiency as well. In the short run, however, there are a number of issues related to transparency, investors' misplaced trust, and poorly aligned incentives between e-brokerages and market makers, which may impede true market efficiency.

For efficiency to move beyond the user interface and into the trading process, individual investors need a transparent window to observe the actual flow of orders, the time of execution, and the commission structure at various points in the trading process. In this regard, institutional rules, regulations, and monitoring functions play a significant role in promoting efficiency and transparency along the value chain in online trading markets.

References

- Barnett, M. 1999a. Day trading can damage your wealth. *Industry Standard* 5: 19.
- . 1999b. Individual investors are becoming institutions. *Industry Standard*, 19–26, 33–34.
- Becker, G. 1998. You want high returns? Brace yourself for high risk. *Business Week* 19: 15.
- Brockman, P., and D. Chung. 2000. Informed and uninformed trading in an electronic, order-driven environment. *Financial Review* 50: 125–146.
- Campbell, J. 1987. Stock returns and the term structure. *Journal of Financial Economics* 18: 373–399.
- . 1996. Understanding risk and return. *Journal of Political Economy* 104: 298–345.

- Campbell, J., and R. Shiller. 1991. Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies* 58: 495–514.
- Cooper, J. 2001. Hey, chicken littles, the sky isn't falling. *Business Week* 12: 43.
- Farrell, C. 2000. Online or off, the rules are the same. *Business Week* 22: 148–149.
- Gimein, M. 1999. Playing the net stock game. *Industry Standard* 25: 20–21.
- Glassman, J., and K. Hassett. 1998. Are stocks overvalued? Not a chance. *Wall Street Journal* 30: A18.
- Gogoi, P. 2000. Rage against online brokers. *Business Week* 20: EB98–EB102.
- McLean, B. 1999. Net stocks have their seasons too. *Fortune* 6: 305–308.
- McNamee, M. 2000. Trading online: It's a jungle out there. *Business Week* 22: 168–169.
- Mintel International Group Ltd. 2002. Online trading market – US report, 1.
- Mullaney, T. 2001. What's an investor to do? *Business Week* 19: EB14–EB15.
- Opiela, N. 2000. Online trading: Opportunity or obstacle? *Journal of Financial Planning* 10: 54–60.
- Pethokoukis, J. 1999a. The young and the fearless. *Business Week* 1: 63–64.
- . 1999b. Forget the cyclone. The net ride is scarier. *US News and World Report* 22: 55.
- Reardon, T. 2000. The price is right. *Accountancy* 55: 81.
- Sharma, M. 2000. The growth of web-based investment. *Information Systems Management* 17: 58–64.
- Smith, G. 1999a. Day of reckoning for day-trading firms? *Business Week* 18: 88–89.
- . 1999b. Valuing those Internet stocks. *Business Week* 8: 87.
- Thornton, E. 2000. Take that, cyber boy. *Business Week* 10: 58–59.
- Vogelstein, F. 1999a. Online traders beware! *US News and World Report* 18: 41–42.
- . 1999b. A virtual stock market. *US News and World Report* 26: 47–48.
- Weber, J. 1999. The market: Too high? Too low? *Business Week* 5: 92–93.

A Critical Evaluation of the Portfolio Performance Indices Under Rank Transformation

22

Ken Hung, Chin W. Yang, Matthew Brigida, and Dwight B. Means

Abstract

This chapter analytically determines the conditions under which four commonly utilized portfolio measures (the Sharpe index, the Treynor index, the Jensen alpha, and the adjusted Jensen alpha) will be similar and different. If the single index CAPM model is appropriate, we prove theoretically that well-diversified portfolios must have similar rankings for the Treynor, Sharpe indices, and adjusted Jensen's alpha ranking. The Jensen alpha rankings will coincide if and only if the portfolios have similar betas. For multi-index CAPM models however, the Jensen alpha will not give the same ranking as the Treynor index even for portfolios of large size and similar betas. Furthermore, the adjusted Jensen's alpha ranking will not be identical to the Treynor index ranking.

Keywords

Sharpe index · Treynor index · Jensen alpha · Adjusted Jensen alpha · CAPM model · Multi-index CAPM model · Performance measures · Rank correlation · Ranking · Rank transformation

K. Hung (✉)

Division of International Banking and Finance Studies, Texas A&M International University, Laredo, TX, USA

e-mail: ken.hung@tamiu.edu

C. W. Yang

Clarion University of Pennsylvania, Clarion, PA, USA

National Chung Cheng University, Chia-yi, Taiwan

e-mail: yang@clarion.edu; yang@mail.clarion.edu; ecdyew@ccu.edu.tw

M. Brigida

SUNY Polytechnic Institute, Utica, NY, USA

e-mail: bridgm@sunypoly.edu

D. B. Means

Clarion University of Pennsylvania, Clarion, PA, USA

22.1 Introduction

Measurement of a portfolio's performance is of extreme importance to investment managers. That is, if a portfolio's risk-adjusted rate of return exceeds (or is below) that of a randomly chosen portfolio, it may be said that it outperforms (or underperforms) the market. The risk-return relation can be dated back to Tobin (1958), Markowitz (1959), Sharpe (1964), Litner (1965), and Mossin (1966). Evaluation measures are attributed to Treynor (1965), Sharpe (1966), and Jensen (1968, 1969). Empirical studies of these indices can be found in the work by Friend and Blume (1970), Black, Jensen, and Scholes (1972), Klemkosky (1973), Fama and MacBeth (1973), and Kim (1978). For instance, the rank correlation between the Sharpe and Treynor indices was found by Sharpe (1966) to be 0.94. Reilly (1986) found the rank correlation to be 1 between the Treynor and Sharpe indices; 0.975 between the Treynor index and Jensen alpha; and 0.975 between the Sharpe index and Jensen alpha. In addition, the sampling properties and other statistical issues of these indices have been carefully studied by Levy (1972), Johnson and Burgess (1975), Burgess and Johnson (1976), Lee (1976), Levhari and Levy (1977), Lee and Jen (1978), and Chen and Lee (1981, 1984, 1986). For example, Chen and Lee (1981, 1986) found that the statistical relationship between performance measures and their risk proxies would, in general, be affected by the sample size, investment horizon, and market conditions associated with the sample period. Notwithstanding these empirical findings, an analytical study of the relationship among these measures is missing in the literature. These performance measures may well be considered very "similar" owing to the unusually high rank correlation coefficients in the empirical studies. However, the empirical findings do not prove the true relationship. These measures can theoretically yield rather divergent rankings especially for the portfolios whose sizes are substantially less than the market. A portfolio size about 15 or more in which further decreases in risk is in general not possible (Evans and Archer 1968; Wagner and Lau 1971; Johnson and Shannon 1974) can generate rather different rankings. In the case of an augmented CAPM, a majority of these performance measures, contrary to the conventional wisdom, can be rather different regardless of the portfolio sizes! In this note, it is our intention to (1) investigate such relationship, (2) clarify some confusing issues, and (3) provide some explanations as to the empirically observed high rank correlations among performance measures. The analysis is free from the statistical assumptions (e.g., normality) and may provide some guidance to portfolio managers.

22.2 The Relationship Between Treynor, Sharpe, and Jensen's Measures in the Simple CAPM

Given the conventional assumptions, a typical CAPM formulation can be shown as

$$y_i = \alpha_i + \beta_i x \quad (1)$$

where $y_i = \pi_p - \pi_f$ = the estimated excess rate of return of portfolio i over the risk-free rate

$x = \pi_m - \pi_f$ = the excess rate of return of the market over the risk-free rate

The Treynor index is a performance measure which is expressed as the ratio of the average excess rate of return of a portfolio over the estimated beta or

$$T_i = \frac{\bar{y}_i}{\beta_i} \quad (2)$$

Similarly, the Sharpe index is the ratio of the average excess rate of return of a portfolio over its corresponding standard deviation or

$$S_i = \frac{\bar{y}_i}{S y_i} \quad (3)$$

A standard deviation which is significantly larger than the beta may be consistent with the lack of complete diversification. While the Sharpe index uses the total risk as denominator, the Treynor index uses only the systematic risk or estimated beta. Note that these two indices are relative performance measures, i.e., relative rankings of various portfolios. Hence, they are suitable for a nonparametric statistical analysis such as rank correlation.

In contrast to these two indices, the Jensen alpha (or α) can be tested parametrically by the conventional t statistic for a given significance level. However, the absolute Jensen alpha may not reflect the proper risk adjustment level for a given performance level (Francis 1980). For instance, two portfolios with the identical Jensen's alpha may well have different betas. In this case, the portfolio with lower beta is preferred to the one with higher beta. Hence, the adjusted Jensen alpha can be formulated as the ratio of the Jensen alpha divided by its corresponding beta (see Francis 1980) or

$$AJ_i = \frac{\alpha_i}{\beta_i} \quad (4)$$

The close correlation between the Treynor and Sharpe indices is often cited in the empirical work of mutual fund performances. Despite its popular acceptance, it is appropriate to examine them analytically by increasing the portfolio size (n) to the number of securities of the market (N), i.e., the portfolio risk premium x approaches the market risk premium y . Rewriting the Treynor index, we have

$$T_i = \frac{y_i}{\beta_i} = y_i$$

$$\frac{y_i}{S y_i} \left[\frac{Var(x)}{S y_i} \right] = S_i \cdot s_x \quad (5)$$

since $Cov(x \cdot y_i) = Var(y_i) = Var(x)$ for $x = y_i$.

Equation (5) indicates that the Treynor index, in general, will not equal the Sharpe index even in the case of a complete diversification, i.e., $n = N$. It is evident from (5) that these two indices are identical only for $s_x = 1$, a highly unlikely scenario. Since neither the Treynor nor Sharpe index is likely to be normally distributed, a rank correlation is typically computed to reflect their association. Taking rank on both sides of Eq. (5) yields

$$\text{Rank}(T_i) = \text{Rank}(S||i) \cdot s_x \quad (6)$$

since s_x in a given period and for a given market is constant. As a result, the Treynor and the Sharpe indices (which must be different values) give identical ranking as the portfolio size approaches the market size as stated in the following proposition.

Proposition #1:

In a given period and for a given market characterized by the simple CAPM, the Treynor and Sharpe indices give exactly the same ranking on portfolios as the portfolio size (n) approaches the market size (N).

This proposition explains high rank correlation coefficients observed in empirical studies between these indices. Similarly, Eq. (5) also indicates that parametric (or Pearson product) correlation between the Treynor and Sharpe indices approaches 1 as n approaches N for a constant s_x , i.e., T_i is a nonnegative linear transformation of S_i from the origin. In general, these two indices give similar rankings but may not be identical.

The Jensen alpha can be derived from the CAPM for portfolio i :

$$J_i = \alpha_i = \bar{y}_i - \beta_i \bar{x} \quad (7)$$

It can be seen from Eq. (7) that as $n \rightarrow N$, y_i approaches x , and $\beta_i \rightarrow 1$. Hence α_i approaches zero. The relationship of the rankings between the Jensen alpha and the Treynor index ranking are equal can be proved as β_i approaches 1 because:

$$\begin{aligned} \text{Rank}(J_i) &= \text{Rank}(\alpha_i) = \text{Rank}\left(\frac{\alpha_i}{\beta_i}\right) \\ &= \text{Rank}(\bar{y}_i) - \text{Rank}(\beta||i\bar{x}) \\ &= \text{Rank}\left(\frac{\alpha_i}{\beta_i}\right) - \text{Rank}(\bar{x}) \\ &= \text{Rank}(T_i) \end{aligned} \quad (8)$$

Since $\bar{x}$ is a constant; $y_i/\beta_i \rightarrow y_i$ and $\beta_i \bar{x} \rightarrow \bar{x}$. We state this relationship in the following proposition.

Proposition #2:

In a given period and for a given market characterized by the simple CAPM, as the portfolio size n approaches the market size N , the Jensen alpha ranking approaches the Treynor index ranking.

However, the Jensen alpha will in general be dependent on the average risk premium for a given beta value for all portfolios since

$$\begin{aligned}\text{Rank}(\alpha_i) &= \text{Rank}(\bar{y}_i) - \beta_i \text{Rank}(\bar{x}) \\ \text{Rank}(\bar{y}_i) &= \text{constant}\end{aligned}\quad (9)$$

for a constant β_i (for all i). In this case the Jensen alpha will give similar rank to the Treynor index for a set of portfolios with similar beta values since

$$\text{Rank}\left(\frac{y_i}{\beta_i}\right) = \text{Rank}(\bar{y}_i) = \text{Rank}(\alpha_i)$$

for a fairly constant set of β_i 's. Hence, we state the following proposition.

Proposition #3:

In a given period and for a given market characterized by the simple CAPM, the ranking of the Jensen alpha and that of the Treynor index gives very close rankings for a set of fairly similar portfolio betas regardless of the portfolio size.

Next, we examine the relationship between the adjusted Jensen alpha and the Treynor index in the form of the adjusted Jensen alpha (AJ). Since

$$\alpha_i = \bar{y}_i - \beta_i \bar{x}$$

Hence

$$AJ = \frac{\alpha_i}{\beta_i} = \frac{\bar{y}_i}{\beta_i} - \bar{x} \quad (10)$$

It follows immediately from Eq. (10) that

$$\text{Rank}(AJ) = \text{Rank}(T) - \text{Rank}(\bar{x}) \quad (11)$$

The result is stated in the following proposition.

Proposition #4:

In a given period and for a given market characterized by the simple CAPM, the adjusted Jensen alpha gives precisely identical rankings as does its corresponding Treynor index regardless of the portfolio size.

Clearly, it is the adjusted Jensen alpha that is identical to the Treynor index in evaluating portfolio performances in the framework of the simple CAPM. The confusion of these measures can lead to erroneous conclusions. For example, Radcliffe (1990, p. 209) stated that "the Jensen and Treynor measures can be shown to be virtually identical." Since he used only the Jensen alpha in his text, the statement is not correct without further qualifications such as proposition #3. The ranking of the Jensen alpha must equal that of the adjusted Jensen alpha for a set of similar betas, i.e., $\text{Rank}(\alpha_i/\beta_i) = \text{Rank}(\alpha_i)$ for a constant beta across all i . Another useful performance measure is the information ratio (IR) defined as:

$$IR_i = \frac{\alpha_i}{S_{e_i}} \quad (12)$$

where i denotes a particular portfolio, and S_e is the standard deviation of the error term in an estimated CAPM. An examination immediately reveals that:

$$IR_i = \frac{\alpha_i}{S_{e_i}} = \frac{\bar{y}_i}{S_{e_i}} - \beta_i \frac{\bar{x}_i}{S_{e_i}} \quad (13)$$

and hence:

$$\frac{IR_i}{\beta_i} = \left(\frac{\bar{y}_i}{\beta_i} - \bar{x}_i \right) S_{e_i} \quad (14)$$

Note that β_i , $\bar{x}_i$, and S_{e_i} are constant for a given portfolio and as such:

$$\beta_i^{-1} \text{Rank}(IR_i) = \text{Rank}(T_i) S_{e_i} \quad (15)$$

or

$$\text{Rank}(IR_i) = \text{Rank}(T_i) \beta_i S_{e_i} \quad (16)$$

Proposition #4a:

In a given period and for a given market characterized by the simple CAPM, the rank of the information ratio is the same as that of the Treynor index for a given β_i and S_{e_i} . A note of caution in the information ratio is that $IR_i = \frac{\alpha_i}{S_{e_i}}$ is undefined when $n = N$ or $\alpha_i = S_{e_i} = 0$. We rule out the case of $n = N$ to avoid the pitfall.

All other relationships can be derived by the transitivity property as shown in Table 1. In the next section, we expand our analysis to the augmented CAPM with more than one independent variable.

22.3 The Relationship Between the Treynor, Sharpe, and Jensen Measures in the Augmented CAPM

An augmented CAPM can be formulated without loss of generality, as

$$y_i = \alpha_i + \beta_i x + \sum c_{ij} z_{ij} \quad (17)$$

where z_{ij} is another independent variable and c_{ij} is the corresponding estimated coefficient. For instance, z_{ij} could be a dividend yield variable (see Litzenberger and Ramaswami 1979, 1980, 1982). In this case again, the Treynor and Sharpe indices have the same numerators as in the case of a simple CAPM, i.e., the Treynor index still measures risk premium per systematic risk (or β_i) and the Sharpe index measures the risk premium per total risk or (s_y). However, if the portfolio beta is sensitive to the additional data on z_{ij} due to some statistical problem (e.g., multi

Table 1 Analytical rank correlation between performance measures: simple CAPM

	Sharpe index S_i	Treynor index T_i	Jenson alpha J_i	Adjusted Jenson alpha AJ_i
Sharpe index S_i	1			
Treynor index T_i	Rank (T_i) = Rank (S_i) $\geq S_X$ Identical ranking as $n \geq N$	1		
Jenson alpha J_i	As $n \geq N$ Rank (J_i) $\geq$ Rank (S_i)	Rank (J_i) $\geq$ Rank (T_i) as $n \geq N$ or $\beta \geq 1$ or Rank (J_i) $\geq$ Rank (T_i) for similar β_i 's	1	
Adjusted Jenson alpha AJ_i	As $n \geq N$ Rank (AJ_i) $\geq$ Rank (S_i)	Rank (AJ_i) = Rank (T_i) regardless of the portfolio size	Rank (α_i/β_i) = Rank (α_i) for similar β_i 's	1

collinearity), the Treynor index may be very sensitive due to the instability of the beta even for large portfolios. In this case, the standard deviations of the portfolio returns and portfolio betas may not have consistent rankings. Barring this situation, these two measures will in general give similar rankings for well-diversified portfolios.

However, in the augmented CAPM framework, the Jensen alpha may very well differ from the Treynor index even for a set of similar portfolio betas.

This can be seen from reranking (α_i) as:

$$\text{Rank}(\alpha|i) = \text{Rank}(\bar{y}_i) - \beta_i \text{Rank}(\bar{x}) - \sum_j \text{Rank} \quad (18)$$

It is evident from Eq. (18) that the Jensen alpha does not give same rank as the Treynor index, i.e., $\text{Rank}(\alpha_i) \neq \text{Rank} \bar{y}_i/\beta_i = \text{Rank}(\bar{y}_i)$ for a set of constant portfolio beta β_i 's. This is because $c_{ij}\bar{z}_{ij}$ is no longer constant; they differ for each portfolio selected even for a set of constant β_i 's (hence $\beta_i * \text{Rank}(\bar{x})$) for each portfolio i as stated in the following proposition.

Proposition #5:

In a given period and for a given market characterized by the augmented CAPM, the Jensen alpha in general will not give the same rankings as will the Treynor index, even for a set of similar portfolio betas regardless of the portfolio size.

Last, we demonstrate that the adjusted Jensen alpha is no longer identical to the Treynor index as shown in the following proposition.

Proposition #6:

In a given period and for a given market characterized by the augmented CAPM, the adjusted Jensen alpha is not identical to the Treynor index regardless of the portfolio size.

Table 2 Analytical rank correlation between performance measures: augmented CAPM

	Sharpe index S_i	Treynor index T_i	Jenson alpha J_i	Adjusted Jenson alpha AJ_i
Sharpe index S_i	1			
Treynor index T_i	Rank (T_i) and Rank (S_i) are similar barring severe multicollinearity or an unstable beta	1		
Jenson alpha J_i	Rank (J_i) not equal Rank (S_i)	Rank (J_i) not equal Rank (T_i) even for a similar beta and regardless of the portfolio size	1	
Adjusted Jenson alpha AJ_i	Rank (AJ_i) not equal Rank (S_i)	Rank (AJ_i) not equal Rank (T_i) regardless of the portfolio size	Rank (AJ_i) $\geq$ Rank (J_i) for a set of similar β_i 's	1

We furnish the proof by rewriting Eq. (17) for each portfolio i as:
Since $\alpha_i = \bar{y}_i - \beta_i \bar{x} - \sum_j c_{ij} \bar{z}_{ij}$ implies

$$\frac{\alpha_i}{\beta_i} = \frac{(\bar{y}_i)}{\beta_i} - \beta_i \bar{x} - \sum_j \left(\frac{c_{ij}}{\beta_i} \right) \bar{z}_{ij}$$

We have

$$\text{Rank}(AJ_i) = \text{Rank}(T_i) - \text{Rank}(\bar{x}) - \sum_j \text{Rank}\left(\frac{c_{ij}}{\beta_i}\right) \bar{z}_{ij} \tag{19}$$

It follows immediately that $\text{Rank}(AJ) \neq \text{Rank}(T)$ in general since the last term of Eq. (19) is not likely to be constant for each estimated CAPM regression. It is to be noted that contrary to the case of the simple CAPM, the adjusted Jensen alpha and the Treynor index do not produce identical rankings. Likewise, for a similar set of β_i 's for all i , the rankings of the Jensen and adjusted Jensen alpha are closely related. Note that the property of transitivity, however, does not apply in the augmented CAPM since the pairwise rankings of T_i and J_i or AJ_i do not converge consistently (Table 2) even for large portfolios.

22.4 Conclusion

In this note, we first assume the validity of the single index CAPM. The CAPM remains the foundation of modern portfolio theory despite the challenge from fractal market hypothesis (Peters 1991) and long memory (Lo 1991). However, empirical results have revalidated the efficient market hypothesis and refute others (Coggins 1998).

Within this domain, we have examined analytically the relationship among the four performance indices without explicit statistical assumptions (e.g., normality). The Treynor and Sharpe indices produce similar rankings only for well-diversified portfolios. In its limiting case, as the portfolio size approaches the market size, the ranking of the Sharpe index becomes identical to the ranking of the Treynor index. The Jensen alpha generates very similar rankings as does the Treynor index only for a set of comparable portfolio betas. In general, the Jensen alpha produces different ranking than does the Treynor index. Furthermore, we have shown that the adjusted Jensen alpha has identical rankings to the Treynor index in the simple CAPM. However, in the case of an augmented CAPM with more than one independent variable, we found that (1) the Treynor index may be sensitive to the estimated value of the beta; (2) the Jensen alpha may not give similar rankings to the Treynor index even with a comparable set of portfolio betas; and (3) the adjusted Jensen alpha does not produce same rankings as that of the Treynor index. The potential difference in rankings in the augmented CAPM suggests that portfolio managers must exercise caution in evaluating these performance indices. Given the relationship among these four indices, it may be necessary in general to employ each of them (except the adjusted Jensen alpha and the Treynor index are identical in ranking in the simple CAPM) since they represent different measures to evaluate the performance of portfolio investments.

22.5 Notes

1. We focus our analysis on the theoretical relationship among these indices in the framework of a true characteristic line. There are statistical distributions of the returns (e.g., normal or log normal) from which the biases of these indices are derived and other statistical issues are discussed in detail by Chen and Lee (1981, 1986). We shall limit our analysis to a pure theoretical scenario where the statistical assumptions are not essential to our analysis. It is to be pointed out that the normality assumption of stock returns in general has not been validated in the literature.
2. This condition is guaranteed if the portfolio y_i is identical to the market (x) or if n is equal to N . In this special case, if the portfolio is weighted according to market value weights, the portfolio is identical to the market so $Cov(x, y_i) = Var(y_i) = Var(x)$.

References

- Black, F., M.C. Jensen, and M. Scholes. 1972. The capital asset pricing model: Some empirical tests. In *Studies in the theory of capital markets*, ed. M.C. Jensen. New York: Praeger Publishers, Inc.
- Burgess, R.C., and K.H. Johnson. 1976. The effects of sampling fluctuations on required inputs of security analysis. *Journal of Financial and Quantitative Analysis* 11: 847–854.

- Chen, S.N., and C.F. Lee. 1981. The sampling relationship between Sharpe's performance measure and its risk proxy: Sample size, investment horizon and market conditions. *Management Science* (June): 607–618.
- . 1984. On measurement errors and ranking of three alternative composite performance measures. *Quarterly Review of Economics and Business* 24: 7–17.
- . 1986. The effects of the sample size, the investment horizon and market conditions on the validity of composite performance measures: A generalization. *Management Science* 32 (11): 1410–1421.
- Coggin, T.D. 1998. Long-term memory in equity style index. *Journal of Portfolio Management* 24 (2): 39–46.
- Evans, J., and S.H. Archer. 1968. Diversification and reduction of dispersion: An empirical analysis. *Journal of Finance* (December): 761–767.
- Fama, E.F., and J.D. MacBeth. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 81: 607–636.
- Francis, J.C. 1980. *Investments: Analysis and management*. 3rd ed. New York: McGraw-Hill Book Company.
- Friend, I., and M.E. Blume. 1970. Measurement of portfolio performance under uncertainty. *American Economic Review* 60: 561–575.
- Jensen, M.C. May, 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23 (3): 389–416.
- . 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolio. *Journal of Business* (April): 167–247.
- Johnson, K.H., and R.C. Burgess. 1975. The effects of sample sizes on the accuracy of E-V and SSD efficient criteria. *Journal of Financial and Quantitative Analysis* 10: 813–848.
- Johnson, K.H., and D.S. Shannon. 1974. A note on diversification and reduction of dispersion. *Journal of Financial Economics* (December): 365–372.
- Kim, T. 1978. An assessment of performance of mutual fund management. *Journal of Financial and Quantitative Analysis* 13 (3): 385–406.
- Klemkosky, R.C. 1973. The bias in composite performance measures. *Journal of Financial and Quantitative Analysis* 8: 505–514.
- Lee, C.F. 1976. Investment horizon and functional form of the capital asset pricing model. *Review of Economics and Statistics* 58: 356–363.
- Lee, C.F., and F.C. Jen. 1978. Effects of measurement errors on systematic risk and performance measure of a portfolio. *Journal of Financial and Quantitative Analysis* 13: 299–312.
- Levhari, P., and H. Levy. 1977. The capital asset pricing model and investment horizon. *Review of Economics and Statistics* 59: 92–104.
- Levy, H. 1972. Portfolio performance and investment horizon. *Management Science* 18 (12): B645–B653.
- Litner, J. 1965. The valuation of risk assets and the selection of risky investment in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–47.
- Litzenberger, R.H., and K. Ramaswami. 1979. The effects of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics* 7 (2): 163.
- . 1980. Dividends, short selling restriction, tax induced investor clienteles and market equilibrium. *Journal of Finance* 35 (2): 469.
- . 1982. The effects of dividends on common stock prices tax effects or information effect. *The Journal of Finance* XXXVII (2): 429.
- Lo, A.W. 1991. Long-term memory in stock market prices. *Econometrica* 59 (5): 1279–1313.
- Markowitz, H. 1959. *Portfolio selection*, Cowles monograph. Vol. 16. New York: Wiley. Chapter 14.
- Mossin, J. 1966. Equilibrium in a capital market. *Econometrica* 34: 768–783.
- Peters, E.E. 1991. *Chaos and order in the capital markets: A new view of cycles, prices and market volatility*. New York: Wiley.

- Radcliffe, R.C. 1990. *Investment: Concepts, analysis, and strategy*. 3rd ed. Glenview: Scott, Foresman and Company.
- Reilly, F.K. 1986. *Investments*. 2nd ed. Chicago: The Dryden Press.
- Sharpe, W. 1964. Capital asset price: A theory of market equilibrium under conditions of risk. *The Journal of Finance* (September): 425–442.
- Sharpe, W.F. 1966. Mutual fund performance. *Journal of Business Supplement on Security Prices* (January): 119–138.
- Tobin, J. 1958. Liquidity preference as behavior toward risk. *The Review of Economic Studies* 26 (1): 65–86.
- Treynor, J.L. 1965. How to rate management of investment funds. *Harvard Business Review* (January–February): 63–75.
- Wagner, W.H., and S. Lau. 1971. The effect of diversification on risk. *Financial Analysts Journal* (November–December): 48–53.

Corporate Failure: Definitions, Methods, and Failure Prediction Models

23

Jenifer Piesse, Cheng-Few Lee, Hsien-Chang Kuo, and Lin Lin

Abstract

The exposure of a number of serious financial frauds in high-performing listed companies during the past couple of years has motivated investors to move their funds to more reputable accounting firms and investment institutions. Clearly, bankruptcy, or corporate failure or insolvency, resulting in huge losses has made investors wary of the lack of transparency and the increased risk of financial loss. This chapter provides definitions of terms related to bankruptcy and describes common models of bankruptcy prediction that may allay the fears of investors and reduce uncertainty. In particular, it will show that a firm filing for corporate insolvency does not necessarily mean a failure to pay off its financial obligations when they mature. An appropriate risk-monitoring system, based on well-developed failure prediction models, is crucial to several parties in the investment community to ensure a sound financial future for clients and firms alike.

Keywords

Bankruptcy · Conditional Probability Analysis (CPA) · Corporate failure · Discriminant Analysis (DA) · Distress · Failure prediction · Hazard models · Liquidation · Misclassification cost models · Receivership

J. Piesse

University of London, London, UK

e-mail: jenifer.piesse@kcl.ac.uk

C.-F. Lee (✉)

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: lee@rbs.rutgers.edu

H.-C. Kuo · L. Lin

National Chi-Nan University, Nantou, Taiwan, Republic of China

23.1 Introduction

The financial stability of firms is of concern to many agents in society, including investors, bankers, governmental and regulatory bodies, and auditors. The credit rating of listed firms is an important indicator, both to the stock market for investors to adjust stock portfolios, and also to the capital market for lenders to calculate the costs of loan default and borrowing conditions for their clients. It is also the duty of government and the regulatory authorities to monitor the general financial status of firms in order to make proper economic and industrial policy. Further, auditors need to scrutinize the going-concern status of their clients to present an accurate statement of their financial standing. The failure of one firm can have an effect on a number of stakeholders, including shareholders, debtors, and employees. However, if a number of firms simultaneously face financial failure, this can have a wide-ranging effect on the national economy and possibly on that of other countries. A recent example is the financial crisis that began in Thailand in July 1997, which affected most of the other Asia-Pacific countries. For these reasons, the development of theoretical bankruptcy prediction models, which can protect the market from unnecessary losses, is essential. Using these, governments are able to develop policies in time to maintain industrial cohesion and minimize the damage caused to the economy as a whole.

Several terms can be used to describe firms that appear to be in a fragile financial state. From standard textbooks, such as Brealey et al. (2001) and Ross et al. (2002), definitions are given of distress, bankruptcy, or corporate failure. Pastena and Ruland (1986, p. 289) describe this condition as when

1. The market value of assets of the firm is less than its total liabilities;
2. The firm is unable to pay debts when they come due;
3. The firm continues trading under court protection.

Of these, insolvency, or the inability to pay debts when they are due, has been the main concern in the majority of the early bankruptcy literature. This is because insolvency can be explicitly identified and also serves as a legal and normative definition of the term “bankruptcy” in many developed countries. However, the first definition is more complicated and subjective in the light of the different accounting treatments of asset valuation. Firstly, these can give a range of market values to the company’s assets and second, legislation providing protection for vulnerable firms varies between countries.

23.2 The Possible Causes of Bankruptcy

Insolvency problems can result from endogenous decisions taken within the company or a change in the economic environment, essentially exogenous factors. Some of the most common causes of insolvency are suggested by Rees (1990):

- Low and declining real profitability
- Inappropriate diversification: moving into unfamiliar industries or failing to move away from declining ones
- Import penetration into the firm's home markets
- Deteriorating financial structures
- Difficulties controlling new or geographically dispersed operations
- Over-trading in relation to the capital base
- Inadequate financial control over contracts
- Inadequate control over working capital
- Failure to eliminate actual or potential loss-making activities
- Adverse changes in contractual arrangements.

Apart from these, a new company, is usually thought to be riskier than those with longer history. Blum (1974, p. 7) confirmed that “other things being equal, younger firms are more likely to fail than older firms.” Hudson (1987), examining a sample between 1978 and 1981, also pointed out that companies liquidated through a procedure of creditors' voluntary liquidation or compulsory liquidation during that period were on average 2–4 years old and three-quarters of them less than 10 years old. Moreover, Walker (1992, p. 9) also found that “many new companies fail within the first 3 years of their existence.” This evidence suggests that the distribution of the failure likelihood against the company's age is positively skewed. However, a clear-cut point in age structure has so far not been identified to distinguish “new” from “young” firms in a business context, nor is there any convincing evidence with respect to the propensity to fail by firms of different ages. Consequently, the age characteristics of liquidated companies can only be treated as an observation rather than theory.

However, although the most common causes of bankruptcy can be noted, they are not sufficient to explain or predict corporate failure. A company with any one or more of these characteristics is not certain to fail in a given period of time. This is because factors such as government intervention may play an important role in the rescue of distressed firms. Therefore, as Bulow and Shoven (1978) noted, the conditions under which a firm goes through liquidation are rather complicated. Foster (1986, p. 535) described this as “there need not be a one-to-one correspondence between the non-distressed/distressed categories and the non-bankrupt/bankrupt categories.” It is noticeable that this ambiguity is even more severe in the not-for-profit sector of the economy.

23.3 Methods of Bankruptcy

As corporate failure is not only an issue for company owners and creditors but also the wider economy, many countries legislate for formal bankruptcy procedures for the protection of the public interest, such as ► Chaps. 8, “[Treasury Inflation-Protected Securities](#),” and ► 12, “[Working Capital and Cash Flow](#),” in the US, and the Insolvency Act in the UK. The objective of legislation is to “[firstly] protect the rights of creditors. . . [secondly] provide time for the distressed business to improve

its situation . . . [and finally] provide for the orderly liquidation of assets” (Pastena and Ruland 1986, p. 289). In the UK, where a strong rescue culture prevails, the Insolvency Act contains six separate procedures, which can be applied to different circumstances to prevent either creditors, shareholders, or the firm as a whole from unnecessary loss, thereby reducing the degree of individual as well as social loss. They will be briefly described in the following section.

23.3.1 Company Voluntary Arrangements

A voluntary arrangement is usually submitted by the directors of the firm to an insolvency practitioner, “who is authorised by a recognised professional body or by the Secretary of State” (Rees 1990, p. 394) when urgent liquidity problems have been identified. The company in distress then goes through the financial position in detail with the practitioner and discusses the practicability of a proposal for corporate restructuring. If the practitioner endorses the proposal, it will be put to the company’s creditors in the creditors’ meeting, requiring an approval rate of 75% of attendees. If the restructuring report is accepted, those notified will thus be bound by this agreement and the practitioner becomes the supervisor of the agreement. It is worth emphasizing that a voluntary arrangement need not pay all the creditors in full but a proportion of their lending (30% in a typical voluntary agreement in the UK) on a regular basis for the following several months. The advantages of this procedure are that it is normally much cheaper than formal liquidation proceedings and the creditors usually receive a better return.

23.3.2 Administration Order

It is usually the directors of the insolvent firm who petition the court for an administration order. The court will then assign an administrator, who will be in charge of the daily affairs of the firm. However, before an administrator is appointed, the company must convince the court that the making of an order is crucial to the survival of the company or for a better realization of the company’s assets than would be the case if the firm were declared bankrupt. Once it is rationalized, the claims of all creditors are effectively frozen. The administrator will then submit recovery proposals to the creditors’ meeting for approval within 3 months of the appointment being made. If this proposal is accepted, the administrator will then take the necessary steps to put it into practice.

An administration order can be seen as the UK version of the US ► [Chap. 12, “Working Capital and Cash Flow,”](#) in terms of the provision of a temporary legal shelter for troubled companies. In this way, they can escape future failure without damaging their capacity to continue to trade (Counsell 1989). This does sometimes lead to insolvency avoidance altogether (Homan 1989).

23.3.3 Administrative Receivership

An administration receiver has very similar powers and functions as an administrator but is appointed by the debenture holder (the bank), secured by a floating or fixed charge after the directors of the insolvent company see no prospect of improving their ability to honor their debts. In some cases, before the appointment of an administration receiver, a group of investigating accountants will be empowered to examine the real state of the company. The investigation normally includes the estimation of the valuable assets and liabilities of the company. If this group finds that the company has no other choices but to be liquidated, an administration receiver will work in partnership with the investigation team and thus be entitled to take over the management of the company. The principal aim is to raise money to pay debenture holders and other preferential creditors by selling the assets of the businesses at the best price. The whole business may be sold as a going concern if it is worth more as an entity. As in an administration order, the receiver must advise creditors of any progress through a creditors' meeting, which is convened shortly after the initial appointment.

23.3.4 Creditors' Voluntary Liquidation

In a creditors' voluntary liquidation, the directors of the company will take the initiative to send an insolvency practitioner an instruction that will lead to the convening of a creditors' and shareholders' meetings. In a shareholders' meeting, a liquidator will be appointed and this is ratified in a subsequent creditors' meeting. Creditors have the right to determine who acts as liquidator. A liquidator will start to find potential purchasers and realise the assets of the insolvent firm in order to clear its debts. Unlike receivers who have wide ranging powers in the management of the businesses, the liquidator's ability to continue trading is restricted. This is the most common way to terminate a company (Rees 1990).

23.3.5 Members' Voluntary Liquidation

The procedure for a member's voluntary liquidation is similar to that of the creditors' voluntary liquidation. The only difference is that in a members' voluntary liquidation the directors of the firm must swear a declaration of solvency to clear debts with fair interest within 12 months and creditors are not involved in the appointment of a liquidator. Therefore, a company's announcement of a members' voluntary liquidation by no means signals its insolvency, but only means closure with diminishing activity, purely a necessity to remain in existence.

23.3.6 Compulsory Liquidation

A compulsory liquidation is ordered by the court to wind up a company directly. This order is usually initiated by the directors of the insolvent firm or its major creditors. Other possible petitioners include the Customs and Excise, the Inland Revenue, and local government (Hudson 1987, p. 213). The entire procedure is usually started with a statutory demand made by creditors who wish to initiate a compulsory liquidation. If the firm fails to satisfy their request in a stated period of time, this failure is sufficient grounds to petition the court to wind up the firm. Once the order is granted, the Official Receiver will take control of the company immediately or a liquidator will be appointed by the Official Receiver. The company then must cease trading and liquidation of assets begins. However, an interesting phenomenon is that many valuable assets may be removed or sold prior to the liquidator taking control, or even during the delivery of the petition to the court, leaving nothing valuable for the liquidator to deal with. In this sense, the company initiating a compulsory liquidation has been terminated in practical terms far before a court order is granted.

23.4 Prediction Model for Corporate Failure

Because corporate failure is not simply the closure of a company but has wider implications, it is important to construct models of corporate failure for assessment and prediction. If bankruptcy can be predicted accurately, it may be possible for the firm to be restructured, thus avoiding failure. This would benefit owners, employees, creditors, and shareholders alike.

There is an established literature that supports the prediction of corporate failure using financial ratio analysis. This is because by using financial performance data it is possible to control for the systematic effect of firm size and industry effects (Lev and Sunder 1979, pp. 187–188) in cross-section models to determine if there are signs of corporate failure. Thus, there is a history of financial ratio analysis in bankruptcy prediction research.

23.4.1 Financial Ratio Analysis and Discriminant Analysis

The earliest example of ratio analysis in predicting corporate failure is attributed to FitzPatrick (1932), although it attracted more attention with the univariate studies of Beaver (1966). This work systematically categorized 30 popular ratios into six groups, and found that some ratios, such as cash flow/total debt ratio, demonstrated excellent predictive power in corporate failure models. These results also showed the deterioration of the distressed firms prior to failure, including a fall in net income, cash flow, and working capital, as well as an increase in total debt. Although this was a useful beginning, univariate analysis was later found to be limited and better results were obtained from including a number of ratios that combined to give a more robust model with improved predictive power.

With the increased popularity of the multi-ratio analysis, multivariate discriminant analysis (MDA) began to dominate the bankruptcy prediction literature from the 1980s. MDA determines the discriminant coefficient of each of the characteristics chosen in the model on the basis that these will discriminate efficiently between failed and nonfailed firms. A single score for each firm in the study is generated and a cut-off point determined that minimizes the dispersion of scores associated with firms in each category, including the probability of overlap between them. An intuitive advantage of MDA is that the model considers the entire profile of characteristics and their interaction. Another advantage lies in its convenience in application and interpretation (Altman 1983, pp. 102–103).

One of the most popular MDA applications is the Z-score model developed by Altman (1968). Because of the success of the Z-score in predicting failure, 22 selected financial ratios were classified into five bankruptcy-related categories. In a sample of 33 bankrupt and 33 nonbankrupt manufacturing companies between 1946 and 1965, the final specification model determined the five variables, which are still frequently used in the banking and business sectors. The linear function is

$$Z\text{-score} = 1.2Z_1 + 1.4Z_2 + 3.3Z_3 + 0.6Z_4 + 0.999Z_5 \quad (1)$$

where

Z-score = overall index;

Z_1 = working capital/total assets;

Z_2 = retained earnings/total assets;

Z_3 = earnings before interest and taxes/total assets;

Z_4 = market value of equity/book value of total debt;

Z_5 = sales/total assets.

Altman (1968) also tested the cut-off point to balance Type I and Type II errors, and found that in general, it was possible for a company with a Z-score smaller than 1.8 to fail during the next few years whereas one with a Z-score higher than 2.99 was much more likely to succeed. The Z-score model remains popular as an indicator of credit risk for banks and other lenders.

Although these statistical discrimination techniques are popular in predicting bankruptcy, there are a number of methodological problems associated with them. Some are a function of the properties of financial ratios, for example, proportionality and zero-intercept assumptions are both critical to the credibility of the ratio analysis. The basic ratio form is assumed to be $y/x = c$, where y and x are two accounting variables that are different but linearly related and c is the value of the ratio. This raises three questions. First, is there an error term in the relationship between the two accounting variables? Second, is an intercept term likely to exist in this relationship? And finally, supposing the numerator and denominator are not linearly related?

With respect to the first question, Lev and Sunder (1979) proved that if there is an additive error term in the relationship between y and x suggested by the underlying theory, that is, $y = \beta x + e$ or $y/x = \beta + e/x$, the comparability of such ratios will be

limited. This is because “the extent of deviation from perfect size control depends on the properties of the error term and its relation to the size variable, x ” (Lev and Sunder 1979, p. 191). The logic is as follows: Where the error term is homoscedastic, e/x is smaller for large firms than for small ones because x as a size variable for large firms will, on average, be greater than that of small firms. Therefore, the ratio y/x for large firms will be closer to the slope term β than that for small firms. Then, since the variance of the ratio y/x for smaller firms is greater than that of larger firms, it proves that the ratio y/x of two groups (large and small firms) are statistically drawn from two different distributions. This weakens the validity of the comparison between ratios. Furthermore, to include an additive error term in the relationship between the numerator and the denominator is not adequate as a size control.

However, if y is heteroscedastic, it may result in the homoscedasticity of y/x . But it is also possible that this heteroscedastic problem of y/x remains unchanged. Lev and Sunder (1979) note that this problem may be ameliorated only when the error term is multiplicative in the relationship, that is, $y = \beta xe$ or $y/x = \beta e$. This is because the deviation of y/x now has no mathematical relationship with the size variable x . As a result, this form of the ratio is more appropriate for purposes of comparison.

The same argument can be applied where an intercept term exists in the relationship between two ratio variables, represented by $y = \alpha + \beta x$ or $y/x = \beta + \alpha/x$. It is clear that the variance of y/x for smaller firms will be larger than that for larger firms under the influence of the term α/x . Again, this is not appropriate in comparisons of corporate performance.

If two variables are needed to control for the market size of y , such as $y = \alpha + \beta x + \delta z$, or $y = \alpha + \beta x + \delta x^2$ if the underlying relationship is nonlinear, the interpretation of the ratios can be ambiguous. All those problems cast doubt on the appropriateness of ratios in a number of situations. Theoretically, use of ratios is less problematic if and only if highly restrictive assumptions are satisfied. Empirically, Whittington (1980) claimed that violation of the proportionality assumption of the ratio form is the most common problem in research using financial data, especially in a time-series analysis at firm level. McDonald and Morris (1984, p. 96) found that the proportionality assumption is better satisfied when a group of firms in a simple homogeneous industry is analyzed, otherwise some amendment of the form of the ratios will be necessary. However, the replacement of the basic form of the ratio with a more sophisticated one is not a solution. On the contrary, on average, the basic form of the ratio performed quite satisfactorily in empirical studies. Keasey and Watson (1991, p. 90) also suggested that possible violations of the proportionality assumptions can be ignored, and since no further theoretical advances have been made on the topic, basic ratio analysis is still common in bankruptcy research.

In addition to the flaws in the design of financial ratios, there are other methodological problems associated with the use of MDA. Of these, non-normality, inequality of dispersion matrices across all groups, and nonrandom sampling are the most prevalent. The violation of the normality assumption has been extensively discussed in the literature since the 1970s (Kshirsagar 1971; Deakin 1976; Eisenbeis 1977; Amemiya 1981; Frecka and Hopwood 1983; Zavgren 1985; Karels and Prakash 1987). Non-normality results in biased tests of significance and estimated

error rates. Studies on univariate normality of financial ratios found that these distributions tended to be skewed (Deakin 1976; Frecka and Hopwood 1983; Karels and Prakash 1987). If the ratios included in the model are not perfectly univariate normal, their joint distribution will, a priori, not be multivariate normal (Karels and Prakash 1987). Therefore, data used in bankruptcy modeling should seek to minimize multivariate non-normality problems. The traditional stepwise procedure does not satisfy this requirement. However, despite several complementary studies on data transformation and outlier removal for ratio normality (Eisenbeis 1977; Ezzamel et al. 1987; Frecka and Hopwood 1983), this is rarely used in MDA models (Shailer 1989, p. 57). Because all these techniques are imperfect, McLeay (1986) advocated that selecting a better model is more straightforward than the removal of outliers or data transformations.

Given the problems of non-normality, inequality of dispersion matrices across all groups in MDA modeling is trivial by comparison. In theory, the violation of the equal dispersion assumption will affect the appropriate form of the discriminating function. After testing the relationship between the inequality of dispersions and the efficiency of the various forms of classification models, a quadratic classification rule seems to outperform a linear one in terms of the overall probability of misclassification when the variance–covariance matrices of the mutually exclusive populations are not identical (Eisenbeis and Avery 1972; Marks and Dunn 1974; Eisenbeis 1977). More importantly, the larger the difference in dispersion across groups, the more the quadratic form of the discriminating function is recommended.

One of the strict MDA assumptions is random sampling. However, the sampling method used in bankruptcy prediction studies is choice-based, or state-based, sampling which results in an equal or approximately equal draw of observations from each population group. Because corporate failure is not a frequent occurrence (Altman et al. 1977; Wood and Piesse 1988), such sampling technique will cause a relatively lower probability of misclassifying distressed firms as nondistressed (Type I Error) but a higher rate of misclassifying nondistressed firms as distressed (Type II Error) (Lin and Piesse 2004; Kuo et al. 2002; Palepu 1986; Zmijewski 1984). Therefore, the high predictive power of MDA models claimed by many authors appears to be suspect. Zavgren (1985, p. 20) commented that MDA models are “difficult to assess because they play fast and loose with the assumptions of discriminant analysis.” Where there is doubt about the validity of the results of MDA models, a more robust approach such as conditional probability analysis (CPA) is an alternative.

23.4.2 Conditional Probability Analysis

Since the late 1970s, the use of discriminant analysis has been gradually replaced by the CPA. This differs from MDA in that CPA produces the “probability of occurrence of a result, rather than producing a dichotomous analysis of fail/survive as is the norm with basic discriminant techniques” (Rees 1990, p. 418). CPA primarily refers to logit and probit techniques and has been widely used in bankruptcy research

(Keasey and Watson 1987; Martin 1977; Mensah 1983; Ohlson 1980; Peel and Peel 1987; Storey et al. 1987; Zavgren 1985, 1988). The major advantage of CPA is that it does not depend on the assumptions demanded by MDA (Kennedy 1991, 1992). However, logit CPA is not always preferred under all conditions. If the multivariate normality assumption is met, the MDA Maximum Likelihood Estimator (LME) is more asymptotically efficient than MLE logit models. In all other circumstances, the MLE of MDA models may not be consistent, unlike that of logit models (Amemiya 1981; Judge et al. 1985; Lo 1986). However, as the rejection of normality in bankruptcy literature is very common, the logit model is appealing. Empirically, the logit analysis is most robust in the classification of distress.

The most commonly cited example of CPA research in this field is Ohlson (1980). The sample used included 105 bankrupt and 2058 nonbankrupt industrial companies during 1970–1976, contrasting with earlier studies that used equal numbers of bankrupts and nonbankrupts (Altman 1968). The CPA logit analysis results in prediction failure with an accuracy rate of over 92% and included financial ratios to account for company size, capital structure, return on assets, and current liquidity, among others. This model was specified as:

$$Y = -1.3 - 0.4Y_1 + 6.0Y_2 - 1.4Y_3 + 0.1Y_4 - 2.4Y_5 - 1.8Y_6 + 0.3Y_7 - 1.7Y_8 - 0.5Y_9 \quad (2)$$

where:

Y = overall index;

Y_1 = $\log(\text{total assets}/\text{GNP price-level index})$;

Y_2 = total liabilities/total assets;

Y_3 = working capital/total assets;

Y_4 = current liabilities/current assets;

Y_5 = one if total liabilities exceed total assets, zero otherwise;

Y_6 = net income/total assets;

Y_7 = funds provided by operations/total liabilities;

Y_8 = one if net income was negative for the last 2 years, zero otherwise;

Y_9 = change in net income.

It is interesting to note that Ohlson (1980) chose 0.5 as the cut-off point, implicitly assuming a symmetric loss function across the two types of classification errors. The cut-off point was calculated using data beyond the estimation period, although the characteristics of the CPA model, and the large sample size, neutralized any problems (Ohlson 1980, p. 126). It is important to note that while this was a valid approach for cross-section comparisons, it could not be transferred to comparisons across different time periods. With respect to predictive accuracy rates, Ohlson (1980) found that the overall results of the logit models were no obvious improvement on those from the MDA. Hamer (1983) tested the predictive power of MDA and logit CPA, and concluded that both performed comparably in the prediction of business failure for a given data set. However, given the predictive accuracy rates

were overstated in previous MDA papers, mainly due to the use of choice-based sampling, this comparison may be biased and the inferences from them could favor CPA. Apart from this, other factors discussed in this literature question these comparisons, citing differences in the selection of predictors, the firm matching criteria, the lead time, the estimation and test time periods, and the research methodology. Unless these factors are specifically controlled, any claim about the comparative advantages between CPA and MDA in terms of the predictive ability will not be robust.

In conclusion, CPA provides all the benefits of other techniques, including ease of interpretation, but also has none of the strict assumptions demanded by MDA. Thus, CPA can be claimed to be the preferred approach to bankruptcy classification.

23.4.3 Three CPA Models: LP, PM, and LM

Three commonly cited CPA models are: the linear probability model (LP), the probit model (PM), and the logit model (LM). This technique estimates the probability of the occurrence of a result, with the general form of the CPA equation stated as

$$\begin{aligned}\Pr(y = 1) &= F(x, \beta) \\ \Pr(y = 0) &= 1 - F(x, \beta)\end{aligned}\tag{3}$$

In this specification, y is a dichotomous dummy variable which takes the value of 1 if the event occurs and 0 if it does not, and $Pr()$ represents the probability of this event. $F()$ is a function of a regressor vector x coupled with a vector β of parameters to govern the behavior of x on the probability. The problem arises as to what distribution best fits the above equation. Derived from three different distributions, LP, PM, and LM are then chosen to determine the best fit.

LP is a linear regression model, which is simple but has two main problems in application. The first is the heteroscedastic nature of the error term. Recall the form of an ordinary LP, $Y = X'\beta + \varepsilon$, where Y is the probability of an outcome and X is a column of independent variables, β is the parameter vector, and ε is the error term. When an event occurs, $Y = 1$, $\varepsilon = 1 - X'\beta$; but when it does not occur, $Y = 0$, $\varepsilon = (-X'\beta)$. The second error term is not normally distributed, so Feasible General Least Squares Estimation Procedure (FGLS) should be used to correct heteroscedasticity (Greene 1997, p. 87).

A more serious problem is that LP cannot constrain Y to lie between 0 and 1, as a probability should. Amemiya (1981, p. 1486) then suggested the condition that $Y = 1$ if $Y > 1$ and $Y = 0$ if $Y < 0$. But this can produce unrealistic and nonsensical results. Therefore, LP is rarely used and is discarded in the present study.

In the discussion of qualitative response models, there is a lively debate about the comparative benefits of logit and probit models. Although logit models are derived from a logistic density function and probit models from a normal density function, these two distributions are almost identical except that the logistic distribution has thicker tails and a higher central peak (Cramer 1991, p. 15). This means the

probability at each tail and in the middle of the logistic distribution curve will be larger than that of the normal distribution. However, one of the advantages of using logit is its computational simplicity, shown here in the relevant formulae:

$$\begin{aligned} \text{Probit Model : Prob}(Y = 1) &= \int_{-\infty}^{\beta'x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \\ &= \Phi(\beta'x) \end{aligned} \quad (4)$$

$$\begin{aligned} \text{Logit Model : Prob}(Y = 1) &= \frac{\exp(\beta'x)}{1 + \exp(\beta'x)} \\ &= \frac{1}{1 + \exp(-\beta'x)} \end{aligned} \quad (5)$$

where function $\Phi()$ is the standard normal distribution. The mathematical convenience of logit models is one of the reasons for its popularity in practice (Greene 1997, p. 874).

With respect to classification accuracy of CPA models, some comparisons of the results produced from these two models suggest that they are actually indistinguishable where the data are not heavily concentrated in the tails or the center (Amemiya 1981; Cramer 1991; Greene 1997). This finding is consistent with the difference in the shape of the two distributions from which PM and LM are derived. It is also shown that the logit coefficients are approximately $\pi/\sqrt{3} \approx 1.8$ times as large as the probit coefficients, implying that the slopes of each variable are very similar. In other words, “the logit and probit model results are nearly identical” (Greene 1997, p. 878).

The choice of sampling methods is also important in CPA. The common sampling method in the bankruptcy literature is to draw a sample with an approximately equal number of bankrupts and nonbankrupts, usually referred to as the state-based sampling technique, and is an alternative to random sampling. Although econometric estimation usually assumes random sampling, the use of state-based sampling has an intuitive appeal. As far as bankruptcy classification models are concerned, corporate failure is an event with rather low probability. Hence, a random sampling method may result in the inclusion of a very small percentage of bankrupts but a very high percentage of nonbankrupts. Such a sample will not result in efficient estimates in an econometric model (Palepu 1986, p. 6). In contrast, state-based sampling is an “efficient sample design” (Cosslett 1981, p. 56), which can effectively reduce the required sample size without influencing the provision of efficient estimators if an appropriate model and modification procedure are used. Thus, in bankruptcy prediction, the information content of a state-based sample for model estimation is preferred to that of random sampling. A state-based sample using CPA resulted in an understatement of Type I errors but an overstatement of Type II errors (Palepu 1986; Lin and Piesse 2004).

Manski and McFadden (1981) suggested several alternatives that can minimize the problems of state-based sampling. These include the weighted exogenous sampling maximum likelihood estimator (WESMLE) and the modified version by

Cosslett (1981), the nonclassical maximum likelihood estimator (NMLE), and the conditional maximum likelihood estimator (CMLE). They compare and report these estimation procedures, which can be summarized as follows:

- All these estimators are computationally tractable, consistent, and asymptotically normal.
- The weighted estimator and conditional estimator avoid the introduction of nuisance parameters.
- The nonclassical maximum likelihood estimators are strictly more efficient than the others in large samples.
- In the presence of computational constraints, WESMLE and CMLE are the best; otherwise, NMLE is the most desirable.

Thus, by using any of these modifications, the advantages of using state-based sampling technique can be retained, while the disadvantages can be largely removed. The inference from this comparison is that the selection of modification method depends upon two factors: the sample size and the computational complexity. The modification cited in the bankruptcy literature is CMLE for three main reasons. Firstly, it has been extensively demonstrated in logit studies by Cosslett (1981) and Maddala (1983). Secondly, it was the model of choice in the acquisition prediction model by Palepu (1986), the merger/insolvency model by BarNiv and Hathorn (1997), and the bankruptcy classification models by Lin and Piesse (2004). Finally, because CMLE only introduces a change to the constant term that normally results from MLE estimation, while having no effects on the other parameters, this procedure is relatively simple. Without bias caused by the choice of sampling methods, modified CPA can correct all the methodological flaws of MDA.

23.5 The Selection of an Optimal Cut-Off Point

The final issue with respect to the accuracy rate of a bankruptcy classification model is the selection of an optimal cut-off point. Palepu (1986) noted that traditionally the cut-off point determined in most early papers was arbitrary, usually 0.5. This choice may be intuitive, but lacks theoretical justification. Joy and Tollefson (1975), Altman and Eisenbeis (1978), and Altman et al. (1977) calculated the optimal cut-off point in the ZETA model. Two elements in the calculation can be identified, the costs of Type I and Type II errors and the prior probability of failure and survival, both of which had been ignored in previous studies. However, Kuo et al. (2002) uses fuzzy theory methods to improve a credit decision model.

Although their efforts were important, unsolved problems remain. The first is the subjectivity in determining the costs of Type I and Type II errors. Altman et al. (1977, p. 46) claimed that bank loan decisions will be approximately 35 times more costly when Type I errors occurred than for Type II errors. This figure is specific to the study and is not readily transferred and therefore a more general rule is required.

The second problem is the subjectivity of selecting a prior bankruptcy probability. Wood and Piesse (1988) criticized Altman et al. (1977) for choosing a 2% higher failure rate than the annual average failure rate of 0.5%, suggesting spurious results from Altman et al. and necessitating a correction that was taken up in later research. The final problem is that the optimal cut-off score produced may not be “optimal” when multinormality and equal dispersion matrices assumptions are violated, which is a common methodological problem in this data analysis (Altman et al. 1977, p. 43, footnote 17).

The optimal cut-off equation in Maddala (1983, p. 80) is less problematic. It begins by developing an overall misclassification cost model:

$$C = C_1 P_1 \int_{G_2} f_1(x) dx + C_2 P_2 \int_{G_1} f_2(x) dx \quad (6)$$

where

C = the total cost of misclassification;

C_1 = the cost of mis-classifying a failed firm as a non-failed one (Type I error);

C_2 = the cost of mis-classifying a non-failed firm as a failed one (Type II error);

P_1 = the proportion of the failed firms to the total population;

P_2 = the proportion of the non-failed firms to the total population;

G_1 = the failed firm group;

G_2 = the non-failed firm group;

x = a vector of characteristics $x = (x_1, x_2, \dots, x_k)$;

$f_1(x)$ = the joint distribution of the characteristics x in the failed group;

$f_2(x)$ = the joint distribution of x in the non-failed group.

$P_1 + P_2 = 1$

However,

$$\text{Given } \int_{G_2} f_1(x) dx + \int_{G_1} f_1(x) dx = 1 \quad (7)$$

Combining Eqs. (6) and (7) gives

$$\begin{aligned} C &= C_1 P_1 \left(1 - \int_{G_1} f_1(x) dx \right) + C_2 P_2 \int_{G_1} f_2(x) dx \\ &= C_1 P_1 + \int_{G_1} [C_2 P_2 f_2(x) - C_1 P_1 f_1(x)] dx \end{aligned} \quad (8)$$

then to minimize the total cost of misclassification, $\min C$, it is necessary for

$$C_2 P_2 f_2(x) - C_1 P_1 f_1(x) \leq 0 \quad (9)$$

or

$$\frac{f_1(x)}{f_2(x)} \geq \frac{C_2 P_2}{C_1 P_1} \quad (10)$$

If it is assumed that the expected costs of Type I error and Type II error are equal, $C_2 P_2 = C_1 P_1$, the condition to minimize the total misclassification cost will be

$$\frac{f_1(x)}{f_2(x)} \geq 1 \quad (11)$$

This result is consistent with that proposed by Palepu (1986), assuming equal costs of Type I and II errors. Therefore, the optimal cut-off point is the probability value where the two conditional marginal densities, $f_1(x)$ and $f_2(x)$, are equal. In this equation, there is no need to use the prior failure rate to calculate the optimal cut-off point, the *ex post* failure rate (that is, the sample failure rate). Palepu (1986) illustrates this more clearly using Bayes' theorem.

Instead of using the costs of Type I and Type II errors, the expected costs of these errors are still unknown. Unfortunately, the subjectivity of deciding the relationship between the two types of expected costs still remains. There is no theoretical reason why they should be the same. However, compared to the previous arbitrary 50% cut-off point, this assumption is neutral and therefore preferred. Examples of applications using this method to determine the cut-off probability can be found in Palepu (1986) and Lin and Piesse (2004).

23.6 Recent Developments

While MDA and CPA are classified as static analyses, dynamic modeling is becoming more common in the bankruptcy literature. Shumway (2001) criticized static bankruptcy models for their examination of bankrupt companies 1 year prior to failure, while ignoring changes in the financial status of the firm year to year and proposed a simple dynamic hazard model to assess the probability failure on a continuous basis. Given the historical infrequency of corporate failure, the hazard model avoids the small sample problem because it requires all available time series of firm information. Because the hazard model takes the duration dependence, time-varying covariates, and data sufficiency problems into consideration, it is methodologically superior to both the MDA and CPA family of models. More empirical evidence is needed on its predictive power. Similar studies are in Whalen (1991) and Helwege (1996).

23.7 Conclusion

There are many reasons why a firm may fail and corporate insolvency *does not necessarily* include the inability to pay off financial obligations when they mature. For example, a solvent company can also be wound up through a member's

voluntary liquidation procedure to maximize the shareholders' wealth when the realized value of its assets exceeds its present value in use. Bulow and Shoven (1978) modeled the potential conflicts among the various claimants to the assets and income flows of the company (for example, bondholders, bank lenders, and equity holders) and found that a liquidation decision should be made when "the coalition of claimants with negotiating power can gain from immediate liquidation" (Bulow and Shoven 1978, p. 454). Their model also considered the existence of some asymmetric claims on the firm. This emphasizes the complex nature of bankruptcy decisions and justifies the adoption of members' voluntary liquidation procedure to determine a company's future (see Brealey and Myers 2001, p. 622; Ross and Westerfield 2002, p. 857).

The evolution and development of failure prediction models have produced increasingly superior methods, although an increase of their predictive power does not necessarily correlate with complexity. In addition, the costs of bankruptcy vary with different institutional arrangements and different countries (Brealey and Myers 2001, pp. 439–443; Ross and Westerfield 2002, p. 426). This implies that a single bankruptcy prediction model, with a fixed cut-off probability that can be used for all time periods and in all countries, does not exist. This chapter has raised some of the problems with modeling corporate failure and reviewed some empirical research in the field.

Acknowledgements We would like to thank many friends in University of London (U.K.) and National Chi Nan University (Taiwan) for valuable comments. We also want to thank our research assistant Chiu-mei Huang for preparing the manuscript and proofreading several drafts of the manuscript. Last, but not least, special thanks go to the Executive Editorial Board of the *Encyclopedia in Finance* in Springer, who expertly managed the development process and superbly turned our final manuscript into a finished product.

References

- Altman, E.I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance* 23 (4): 589–609.
- . 1983. *Corporate financial distress: A complete guide to predicting, avoiding, and dealing with bankruptcy*. New York: Wiley.
- Altman, E.I., and R.O. Eisenbeis. 1978. Financial applications of discriminant analysis: A clarification. *Journal of Financial and Quantitative Analysis* 13 (1): 185–195.
- Altman, E.I., R.G. Haldeman, and P. Narayanan. 1977. Zeta analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance* 1: 29–54.
- Amemiya, T. 1981. Qualitative response models: A survey. *Journal of Economic Literature* 19 (4): 1483–1536.
- BarNiv, R., and J. Hathorn. 1997. The merger or insolvency alternative in the insurance industry. *Journal of Risk and Insurance* 64 (1): 89–113.
- Beaver, W. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research* 4 (Suppl): 71–111.
- Blum, M. 1974. Failing company discriminant analysis. *Journal of Accounting Research* 12 (1): 1–25.

- Brealey, R.A., S.C. Myers, and A.J. Marcus. 2001. *Fundamentals of corporate finance*. 3rd ed. New York: McGraw-Hill.
- Bulow, J., and J. Shoven. 1978. The bankruptcy decision. *Bell Journal of Economics* 9 (2): 437–456.
- Cosslett, S.R. 1981. Efficient estimation of discrete-choice models. In *Structural analysis of discrete data with econometric applications*, ed. C.F. Manski and D. McFadden, 51–111. London: MIT Press.
- Counsell, G. 1989. Focus on workings of insolvency act. *The Independent*, April 4.
- Cramer, J.S. 1991. *The logit model: An introduction for economists*. Kent: Edward Arnold.
- Deakin, E.B. 1976. Distributions of financial accounting ratios: Some empirical evidence. *Accounting Review* 51 (1): 90–96.
- Eisenbeis, R.A. 1977. Pitfalls in the application of discriminant analysis in business, finance, and economics. *Journal of Finance* 32 (3): 875–900.
- Eisenbeis, R.A., and R.B. Avery. 1972. *Discriminant analysis and classification procedure: Theory and applications*. Lexington: D.C. Heath and Company.
- Ezzamel, M., M.M. Cecilio, and A. Beecher. 1987. On the distributional properties of financial ratios. *Journal of Business Finance and Accounting* 14 (4): 463–481.
- FitzPatrick, P.J. 1932. A comparison of ratios of successful industrial enterprises with those of failed firms. *Certified Public Accountant*, October, November, and December, 598–605, 656–662, and 727–731, respectively.
- Foster, G. 1986. *Financial statement analysis*. 2nd ed. Hoboken: Prentice-Hall.
- Frecka, T.J., and W.S. Hopwood. 1983. The effects of outliers on the cross-sectional distributional properties of financial ratios. *Accounting Review* 58 (1): 115–128.
- Greene, W.H. 1997. *Econometric analysis*. 3rd ed. Hoboken: Prentice-Hall.
- Hamer, M.M. 1983. Failure prediction: Sensitivity of classification accuracy to alternative statistical methods and variable sets. *Journal of Accounting and Public Policy* 2 (4): 289–307.
- Helwege, J. 1996. Determinants of saving and loan failures: Estimates of a time-varying proportional hazard function. *Journal of Financial Services Research* 10: 373–392.
- Homan, M. 1989. *A study of administrations under the Insolvency Act 1986: The results of administration orders made in 1987*. London: ICAEW.
- Hudson, J. 1987. The age, regional and industrial structure of company liquidations. *Journal of Business Finance and Accounting* 14 (2): 199–213.
- Joy, M.O., and J.O. Tollefson. 1975. On the financial applications of discriminant analysis. *Journal of Financial and Quantitative Analysis* 10 (5): 723–739.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl, and T.C. Lee. 1985. *The theory and practice of econometrics*. New York: Wiley.
- Karels, G.V., and A.J. Prakash. 1987. Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance and Accounting* 14 (4): 573–595.
- Keasey, K., and R. Watson. 1987. Non-financial symptoms and the prediction of small company failure: A test of the Argenti Hypothesis. *Journal of Business Finance and Accounting* 14 (3): 335–354.
- . 1991. Financial distress prediction models: A review of their usefulness. *British Journal of Management* 2 (2): 89–102.
- Kennedy, P. 1991. Comparing classification techniques. *International Journal of Forecasting* 7 (3): 403–406.
- . 1992. *A guide to econometrics*. 3rd ed. Oxford, UK: Blackwell.
- Kshirsagar, A.M. 1971. *Advanced theory of multivariate analysis*. New York: Marcel Dekker.
- Kuo, H.C., S. Wu, L. Wang, and M. Chang. 2002. Contingent fuzzy approach for the development of banks' credit-granting evaluation model. *International Journal of Business* 7 (2): 53–65.
- Lev, B., and S. Sunder. 1979. Methodological issues in the use of financial ratios. *Journal of Accounting and Economics* 1 (3): 187–210.
- Lin, L., and J. Piesse. 2004. The identification of corporate distress in UK industrials: A conditional probability analysis approach. *Applied Financial Economics* 14: 73–82.

- Lo, A.W. 1986. Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *Journal of Econometrics* 31 (3): 151–178.
- Maddala, G.S. 1983. *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.
- Manski, C.F., and D.L. McFadden. 1981. Alternative estimators and sample designs for discrete choice analysis. In *Structural analysis of discrete data with econometric applications*, ed. C.F. Manski and D.L. McFadden, 2–50. London: MIT Press.
- Marks, S., and O.J. Dunn. 1974. Discriminant functions when covariance matrices are unequal. *Journal of the American Statistical Association* 69 (346): 555–559.
- Martin, D. 1977. Early warning of bank failure: A logit regression approach. *Journal of Banking and Finance* 1 (3): 249–276.
- McDonald, B., and M.H. Morris. 1984. The functional specification of financial ratios: An empirical examination. *Accounting and Business Research* 15 (59): 223–228.
- McLeay, S. 1986. Students and the distribution of financial ratios. *Journal of Business Finance and Accounting* 13 (2): 209–222.
- Mensah, Y. 1983. The differential bankruptcy predictive ability of specific price level adjustments: Some empirical evidence. *Accounting Review* 58 (2): 228–246.
- Ohlson, J.A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18 (1): 109–131.
- Palepu, K.G. 1986. Predicting takeover targets: A methodological and empirical analysis. *Journal of Accounting and Economics* 8 (1): 3–35.
- Pastena, V., and W. Ruland. 1986. The merger bankruptcy alternative. *Accounting Review* 61 (2): 288–301.
- Peel, M.J., and D.A. Peel. 1987. Some further empirical evidence on predicting private company failure. *Accounting and Business Research* 18 (69): 57–66.
- Rees, B. 1990. *Financial analysis*. London: Prentice-Hall.
- Ross, S.A., R.W. Westerfield, and J.F. Jaffe. 2002. *Corporate finance*. 6th ed. New York: McGraw-Hill.
- Shailer, G. 1989. The predictability of small enterprise failures: Evidence and issues. *International Small Business Journal* 7 (4): 54–58.
- Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business* 74 (1): 101–124.
- Storey, D.J., K. Keasey, R. Watson, and P. Wynarczyk. 1987. *The performance of small firms*. Bromley: Croom-Helm.
- Walker, I.E. 1992. *Buying a company in trouble: A practical guide*. Hants: Gower.
- Whalen, G. 1991. A proportional hazard model of bank failure: An examination of its usefulness as an early warning tool. *Economic Review, First Quarter* 27: 20–31.
- Whittington, G. 1980. Some basic properties of accounting ratios. *Journal of Business Finance and Accounting* 7 (2): 219–232.
- Wood, D., and J. Piesse. 1988. The information value of failure predictions in credit assessment. *Journal of Banking and Finance* 12: 275–292.
- Zavgren, C.V. 1985. Assessing the vulnerability to failure of American industrial firms: A logistic analysis. *Journal of Business Finance and Accounting* 12 (1): 19–45.
- . 1988. The association between probabilities of bankruptcy and market responses – A test of market anticipation. *Journal of Business Finance and Accounting* 15 (1): 27–45.
- Zmijewski, M.E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* 22 (Suppl): 59–82.



Main Bank Relationships, Debt Structure, and Innovation in Japan

24

Hai-Chin Yu and Phuong Dung Tong

Abstract

This study investigates how a firm's debt structure and the presence of keiretsu main bank relationship conjointly shape a firm's innovation in Japan. Using a 2SLS regression in which corporate debt structure is endogenous, we find that public debt fosters innovation, and bank debt destroys innovation. However, keiretsu main bank relationships can mitigate the negative effect of bank debt. Once a firm is affiliated with a keiretsu, the main bank relationship helps generate 1.57% (or 3.8 more patents) innovation. We explain the plausible reasons are main bank's relationships provide more financial flexibility and risk tolerance to their affiliated firms. Besides, the ownerships of main banks held for their member firms also align the banks' roles with acting more like shareholders than creditors. We identify the causal relation using the event of keiretsu consolidation; the DID (difference-in-differences) results show that the consolidated keiretsu raise more innovation among their member firms by 2.87% after the merger compared to the counterparts. Some further tests show the results are robust.

Keyword

Innovation · Debt structure · Bank debt · Public debt · Main bank

JEL Classification

O31 · O32 · G30

H.-C. Yu

Department of International Business, Chung Yuan University, Chung-Li, Taiwan
e-mail: haichin@cycu.edu.tw

P. D. Tong (✉)

Ministry of Finance, Corporate Finance, Hanoi, Vietnam
e-mail: tongphuongdung@mof.gov.vn

24.1 Introduction

Innovation – technology adaptation and creation – is a crucial engine to sustain firms' competitive advantage and has become increasingly central to firms' growth and economic development (Aghion and Howitt 1992; De Mel et al. 2009). Thus, understanding the determinants of innovation is crucial for policy-makers to conduct an effective policy to boost economic growth.

The extant literature has documented that financial market development relaxes financial constraints and makes it easier for firms to fund investments, including innovative projects. But most focused on the availability of finance, or the choice between equity and debt (Titman and Wessels 1988; Brown et al. 2013; Cornaggia et al. 2015). There is little evidence showing what types of debt financing are more conducive to innovation. This chapter fills the gap by exploring different types of financial arrangements affecting the incentives of firms to innovate and the outcomes.

A growing literature demonstrates that debt financing is critical in spurring innovation by allocating funds for innovation investments. The extant theoretical primary of Aghion and Tilore (1994) links innovation and debt financing and indicates that debts well determine innovation. The empirical work of Ogawa (2007) examines the relationship between debt and innovation input, measured by R&D expenditure, revealing that the debt ratio is significantly negatively associated with innovation in the late 1990s but is insignificant at the end of the 1980s. Recently, Cornaggia et al. (2015) examine bank financing and find that innovation in US small private firms that rely on bank debt has increased following intrastate banking deregulation. Innovation is also found to be associated with bank debt. For example, Chava et al. (2012) document that cheaper bank loans than their peers usually drive firms to produce high-intensive and high-quality patents. The evidence also indicates that bank debt is a common source for corporate innovation (Robb and Robinson 2014), and patents are often used as collateral for bank financing at a certain point (Mann 2018).

By contrast, some research examines public debt and innovation and addresses that short-term bonds discourage innovation projects and reduce investment quality (Hasan et al. 2014). Atanassov (2016) also suggests that firms that rely more on arm's length financing, such as public debt and equity, have larger higher-quality innovation than firms that use other firm sources, such as relationship-based bank financing.

The debate about the fund sources on corporate innovation has caught much attention by scholars. Some research indicates that innovative firms that avoid releasing ongoing creative ideas would use internal funds (Brown et al. 2009). Manso (2011) demonstrates that more innovation outputs are generated when more financial flexibility and risk tolerance are allowed. For example, most IPO firms were initially funded by venture capital (Tian and Wang 2014). Firms that are likely to access external funds generate more patents (Acharya and Xu 2017). In sum, the impacts of fund sources on innovation conclude that external funds with more tolerance and flexibility are suitable for corporate innovation.

Traditional agency theory proposes that banks are risk averse while shareholders are risk taking. When a firm's risk grows, banks become more concerned about the misalignment with the shareholders' interests (Jensen 1986). The potential conflicts between the creditors and shareholders emerged. Although banks are less willing to lend to innovative firms, firms with bank relationships may mitigate this effect. A long-term bank-firm relationship helps create mutual trustworthiness and similar norm; this, in turn, allows banks to give more financial flexibility and risk tolerance to the firms. Bank debt use under these lending relationships may foster innovative activities compared with those without relationships. Recent studies also indicate that banks use more soft information on making a lending decision than hard information (Sumit Agarwal and Hauswald 2010; Saidi and Zaldokas 2021). Since debt is the major funds for Japanese firms (Ogawa 2007), shedding light on debt and innovation is critical for managers, investors, and policy-makers. Again, debts associated with the keiretsu main bank relationships on corporate innovation are still an uncovered field.

We analyze this issue in the setting of Japan, where most of the firms have their main bank relationships that have served as the major creditors for decades. Besides, some large enterprises are affiliated with a close-knit group called a "keiretsu." The core of a keiretsu is usually a large bank (keiretsu bank) that provides capital and loans to group members. These banks also hold stakes in the member firms. In a keiretsu, therefore, banks act both as debtholders and shareholders. The dual nature of the keiretsu main bank relationships changes the traditional structure of the agency conflict between shareholders and debtholders and affects innovation. Therefore, the keiretsu bank relationship is a good setting for exploring debts on firms' innovation.

Besides, Japanese firms rely heavily on financial leverage, in which their debt-to-equity ratios are twice or thrice larger than those of their US counterparts. The higher debt ratio also provides details for exploring the debt structure. Japanese firms are one of the world leaders in the high-technology industry, with more than 50% of firms in the technology-based industry. Under the dual patents system,¹ the sample provides sufficient innovation data for this issue.

Using bank debt and public debt to measure the borrowers' debt structure, we find that, on average, bank debt hinders innovation, whereas public debt use fosters innovation. We find that a 1% increase in bank debt ratio reduced innovation by 1.399%. By contrast, a 1% increase in public debt ratio increased innovation by 4.365%. Although the direct effect of bank debt mitigates corporate innovation, once these firms have a keiretsu main bank relationship, 1.57% more innovation is generated compared with their counterparts that do not have this relationship. These results are robust after a set of endogeneity tests and causal experimental design.

¹ Japanese firms can apply and grant domestic patents under their Japanese names and international patents under their English names. This study uses international patent data granted under the English names to obtain a complete picture of firm innovation. We find the granted number of patents is different from IP-Tech and USPTO. The USPTO shows a lower patent number than the IP-Tech.

We conduct a two-stage least squares (2SLS) analysis using O-scores as an instrument to alleviate the endogeneity concerns of public debt. We also use Japan's keiretsu conglomerate mergers in the 2000s as an experimental design to analyze the causal effect of the keiretsu main bank relationships and innovation. The difference-in-differences (DID) results reveal that the impacts of main bank relationships on innovation raised far more than before after the consolidation.

Our study adds another dimension of keiretsu main bank relationships to the literature of debt structure and innovation. Since innovation financing needs more financial flexibility and risk tolerance, bank debt may not be suitable. However, once bank debt users have keiretsu main bank relationships, the financial restriction of bank debt is released, and the impacts become reverse. We find that the post-consolidation keiretsu banks become more supportive of their member firms and raise the innovation by 2.87%.

The remainder of the study organizes as follows. Section 24.2 reviews the extant literature. Section 24.3 provides the data, sample, and methodology. Sections 24.4 and 24.5 report the empirical results and robustness checks. Section 24.6 concludes.

24.2 Literature Review

24.2.1 The Nature of Innovation and the Roles of Debt on Innovation

The uncertainty of innovation makes it tough to obtain funds. The asymmetric information between investors and innovators makes it difficult to accurately evaluate good projects from the bad ones (Leland and Pyle 1977), then specific innovation projects might be unimplemented. Prior literature states that the trade-off for bank loans in reducing asymmetric information is the holdup costs (Sharpe 1990; Rajan 1992). They argue that the informed banks have more bargaining power to control the investment decision; it continued only when the investment has a positive NPV. However, for a short-term contract, banks tend to reject innovation projects or shut down the ones that are ongoing even it is efficient (Rajan and Zingales 2003).

Besides the mitigation of asymmetric information, some studies indicate that bank loans are reported to foster corporate innovation. Aflonso and Giannangeli (2012) use a large sample of Italian manufacturing SMEs. They conclude that bank debt has a positive impact on raising the probability of outsourcing R&D. Relatedly, Cornaggia et al. (2013) report that after interstate deregulation, the innovation of US small private firms that rely more on bank debt than US public-traded firms increases significantly. Robb and Robinson (2014) indicate that the start-up firms rely heavily on external debt sources; the average start-up firms are about 40% bank financed.

A growing recent literature has documented that debt financing is a vital source for raising external capital for innovation, especially in a bank-based system (Czarnitzki and Kraft 2007). Bank financing is the most common type of debt financing, which was reported to affect the quantity and quality of innovation output

(Backman and Wallin 2018; Brancati 2015; Kerr and Nanda 2015; Nanda and Nicholas 2014; Robb and Robinson 2014).

Even a few works suggest an essential role of bank debt in lending and mitigating asymmetric information; more recent research highlights various constraints that made bank debt unsuitable for financing innovation. Kim and Park (2017) find that the impacts of bank financing on innovation varied among firms due to firms' heterogeneous internal conditions. They conclude that bank financing did not have or have a very trivial effect on innovation for firms with weak incentive systems. Gu et al. (2017) argued that banks are able to affect firms' innovation policies through bankruptcy threat or payment default but thought that bank interventions imposed a negative effect on innovation.

The rational argument behind this is that debt financing helps creditors extract information rents from information monopolies, which negatively influences the quantity and quality of novelty of innovation projects. The negative impacts of debt financing may exist at the first stage of capital seeking, making firms willing to share confidential information on innovation projects with creditors to seek as much capital as possible (Mancusi et al. 2018). However, banks collect the information of the success probabilities and future returns of innovation projects and price the loans above the marginal funding costs to earn positive expected profits (Adam and Streitz 2016; Mattes et al. 2013). The extra rents raise the lending costs and therefore hinder part of the innovative projects.

In comparison, public debt allows managers to pursue innovation as financing flexibility may not be a worry. Bondholders are also not likely to involve in monitoring firms, so managers might make their own decisions in their specialized fields to pursue innovation. Rajan and Zingales (2003) suggest that arm's length financing, such as public debt and equity, is more conducive to creating novel innovations. In contrast, relationship-based bank financing may be better suited to funding less innovative projects or incremental innovations. Atanassov (2016) also documents that firms that rely more on arm's length financing, such as public debt and equity, innovate more and have higher-quality innovations than firms that use other sources, such as bank debt. However, Holmstrom (1989) argues that arm's length capital markets hinder innovation because they pressure management to focus on short-term routine projects at the expense of longer-term innovative projects.

In sum, current literature unequivocally demonstrates the importance of debt financing, but its impact on innovation differs in various contexts. The mixed findings of the effects of bank debt on innovation and the gap for examining the effect of public debt on innovation leave us room to add to the strand of literature.

24.2.2 Main Bank Relationship and Innovation

The feature of Japan's economy is large firms raising capital from banks, and most of them are associated with an organizational structural called Keiretsu. Keiretsu is a set of Japanese banks with interlocking business and shareholding with other companies. Keiretsu banks provide the firms a range of financial services and hold equity

with a quota of less than 5% of the firm's total shares. The shareholding role of Keiretsu reduces the pressures of managers to achieve short-term goals and instead gives them more incentives to pursue long-term goals. Furthermore, each bank has great control over firms since the interlocking role serves to monitor and discipline the firms. The main bank is a tactical financier for the Keiretsu group, to fulfill the double role of lender and shareholder in financing and monitoring.

Another line of literature to explain why the main bank relationship is optimal for firms is reducing information disclosure costs. Borrowing from the main bank not only guarantees sufficient capital but also overcomes leaking information to rivals. In a bank-based financial system, banks and especially main banks are more capable of preserving confidentiality, thus increasing a firm's willingness to disclose information about technology, knowledge, and future business opportunities. Consequently, main banks help to foster firms' innovation and technology adoption. Tadesse (2006) argues that the main bank generated information from leakage, providing more incentives to finance innovation projects.

The prior literature suggests that the close banking relationship is beneficial to firms for improving the availability of bank loans to finance innovation. Hoshi and Kashyap (2001) state that the main bank has close ties to its customers through lending, shareholding, and often has board representation and other personnel placement; main bank relationship is considered to maintain the strong tie with borrowing firms. Fama (1985), Leland and Pyle (1977), Campbell and William (1980), and Diamond (1984, 1991) all point out that close ties banking relationships help raise financing at lower costs. Weinstein and Yafeh (1998) examined the effect of a bank-centered financial system in Japan. They indicated that close bank-firm ties increase the access of capital for innovation. Thus, firms with a close relationship with banks are more likely to access capital easily and invest more in innovation projects.

Rajan (1992) argue that while informed banks make flexible financial decisions that prevent projects from going awry, the cost of this credit is that banks have bargaining power over the firm's profits once projects have begun. Hence, banking relationship makes informed banks willing to lend to the innovative firms. Some scholars also indicated that sharing information between main banks and the borrowing firm is valuable in fostering innovation. King and Levine (1993) suggest that better information-sharing improves the probability of successful innovation. De la Fuente and Marin (1996) show that informed financiers can deter firms' moral hazard by reducing asymmetric information in developing new products. Herrera and Minetti (2007) provide strong empirical evidence of Italian manufacturing firm – where banks represent the main source of external finance – and show that the duration of the banking relationship positively affects the probability that firms innovate. Their argument departs from the theoretical view that information of financiers fosters firms' innovation. A “well-informed financier” can be viewed as the main bank that has incentives and abilities to collect information from the borrowing firms, helping them screen the innovation projects.

Once the bank-firm relationship is established, debt contracts empower incumbent creditors to screen and monitor the ongoing innovation projects and thus

facilitate information collection unavailable for competitor investors (Santos and Winton 2008). Zheng et al. (2018) find that bank-connected firms increase the number of granted patents than the control firms by 18–20% after the monetary stimulus. They benefited from bank financing.

24.3 Data and Methodology

24.3.1 Data

Our patent data obtain from IP Tech – a Japanese patent database website and USPTO. The patent database provides the firms' number of patents annually. The keiretsu main bank data collect from *Industrial Groupings in Japan*. This handbook lists the detailed information of each company and its affiliation to the specific keiretsu. The *Industrial Groupings* reports the Japanese firms' main banks. The main bank holds firms' equity as a shareholder and has seats on board. The close relationship between a firm and the main bank allows a bank to acquire private information from firms and enables firms to access better credit from a bank.

A keiretsu is identified as one of the original eight largest financial institutions, namely, Tokyo-Mitsubishi UFJ, Mitsui, Sumitomo, Dai-Ichi Kangyo, Fuji, Fuyo, Sanwa, and Tokai. As some Japanese banks were hit by increasing bad loans after the 1990s, this situation sparked a keiretsu consolidation in 2000 and 2001. After the keiretsu consolidation in 2001, the leading Japanese keiretsu have Mitsubishi, Mitsui, Sumitomo, Fuyo, Sanwa, and DKB (Mizuho).

The firms' financial statements were obtained from the Global Compustat over 2000–2015. After eliminating financial firms (SIC code ranges from 6000 to 6999) and firms with incomplete data, then merged with IP Tech patent database and USPTO, keiretsu data, and Compustat, our effective observations are 41,299.

24.3.2 Model Specification

The 2SLS regression estimations of Japanese firms' innovation are specified as Eqs. (1) and (2):

$$PUBLICDEBT_{i,t} = \theta_0 + \theta_1 IV + e_{1,t} \quad (1)$$

$$\begin{aligned} INNOVATION_{i,t} = & \beta_0 + \beta_1 BANKDEBT_{i,t} + \beta_2 PUBLICDEBT_{i,t} \\ & + \beta_3 KEIRETSU_{Di,t} + \beta_4 Controls_{i,t} + Industry_i + Year_t \\ & + \varepsilon_{i,t} \end{aligned} \quad (2)$$

INNOVATION is the dependent variable, measured by the natural logarithm of the number of patents that counts for Japanese firms from 2000 to 2015. BANKDEBT is the ratio of bank debt in current liabilities divided by total assets (Low, Glorfeld, Hearth, and Rimbey, 2001). Bank debt comprises bank loans held by a firm,

including bank advances, bank liabilities, bank overdraft, current bank loans, notes payable to banks, and other borrowed funds as reported as part of short-term borrowings. PUBLICDEBT is measured by a firm's bonds and notes in long-term debt, scaled by total assets (Low, Glorfeld, Hearth, and Rimbey, 2001). Public debt details include long-term borrowing and corporate bonds (straight bonds, convertible bonds, and warrant bonds). LEVERAGE is total liabilities scaled by total assets. Main bank keiretsu (KEIRETSU_D) is a dummy variable and equals 1 for firms affiliated to a specific keiretsu and 0 otherwise.

We also control for firm characteristics. LOGTA is the firm size, measured by the logarithm of total assets. AGE is the years since the firm was first incorporated. RD_TA is measured by the ratio of R&D expenditure to total assets. TANGIBLE_TA is the ratio of tangible assets to total assets, in which tangible assets are the net property, plants, and equipment. CAPEX_TA is the capital expenditure ratio, which is capital expenditure divided by total assets. CASH_R is the ratio of cash and cash-equivalent and short-term investments to total assets. EBITDA_TA proxies for the firm performance are calculated by earnings before interests, taxes, and depreciation to total assets. NWC_TA shows the financial health of firms, measured by the amount of working capital scaled by total assets.

IV is an instrument in Eq. (1). Prior research addressed that creditworthiness is crucial for a firm to borrow public financing (Diamond 1991; Hoshi et al. 1993). We thus employ creditworthiness, measured by Ohlson's O-score, as the IV to instruct the firm's public debt use. The O-score is calculated as below:

$$\text{O-SCORE} = -1.32 - 0.407 (\log \text{ of total assets}) + 6.03 (\text{total liabilities/total assets}) - 1.43 (\text{working capital/total assets}) + 0.076 (\text{current liabilities/current assets}) - 1.72 (1 \text{ if total liabilities} > \text{total assets, } 0 \text{ otherwise}) - 0.521 ((\text{net income}_t - \text{net income}_{t-1}) / (|\text{net income}_t| + |\text{net income}_{t-1}|)).$$

24.4 Empirical Results

24.4.1 Summary Statistics

Table 1 presents the summary statistics. Panel A of Table 1 reveals that the median (mean) patents of Japanese firms are 0 (9.5). Japanese firms are highly leveraged with an average leverage ratio 53.7%, which is significantly higher than US firms by approximately 20% during the same period (Lim et al. 2017). Although leverage ratio is high, it is lower than that during 1994 and 2004, which is 59.6% (Yu et al. 2010). Japanese firms have public debt ratio (PUBLICDEBT_TA) of 14.9% on average. The average bank debt ratio (BANKDET_TA) is 37%, indicating that bank debt is more favorable than public debt among Japanese firms. In a bank-based financial system, firms mainly depend on bank loans. Firms with main bank keiretsu relationships (KEIRETSU_D) are 10.22%, indicating a declining trend of keiretsu firms among the listed companies. This ratio is lower than the 19.5% during 1994–2004 (Yu et al. 2010).

Table 1 Summary statistics

Panel A: Summary statistics of full sample							
Variable	Mean	Median	Std.dev	P1	P2	P3	
PATENTS	9.5	0	33.33	0	0	1	
KEIRETSU_D	0.102	0	0.44	0	0	1	
BANKDEBT	0.37	0.376	0.176	0.234	0.505	0.6	
PUBLICDEBT	0.149	0.131	0.101	0.065	0.216	0.294	
LEVERAGE	0.537	0.535	0.225	0.358	0.713	0.85	
TA	276,163	49,443	1,242,224	20,523	148,187	1,082,238	
LOGTA	10.99	10.8	1.524	9.929	11.9	13.894	
AGE	34.13	38	20.06	14	51	64	
RD_TA	0.022	0.014	0.024	0.005	0.03	0.071	
CASH_R	0.167	0.138	0.121	0.08	0.222	0.411	
TANGIBLE_TA	0.28	0.274	0.14	0.18	0.373	0.517	
EBITDA_TA	0.065	0.056	0.048	0.034	0.085	0.151	
CAPEX_TA	0.037	0.029	0.033	0.014	0.05	0.099	
NWC_TA	0.231	0.229	0.182	0.104	0.353	0.54	
O-SCORE	−2.713	−2.711	1.362	−3.692	−1.744	−0.473	
Accrual quality	0.064	0.0161	0.1269	0.0066	0.0577	0.301	
Panel B: Distribution of patents by selected industries from 2000 to 2015							
Industries	Obs.	Mean	p50	p75	p95	p99	std
Agriculture, forestry, and fishing	67	0.2	0	0	1	3	0.5
Chemicals	1870	10.4	1	8	56	118	23.7
Computers	2254	56.8	0	3	179	1815	292.8
Durable manufacturer	9318	24.6	1	6	80	480	150.9
Extractive industries	145	5.8	0	2	35	115	19.3
Foods	829	2.4	0	1	9	62	9.1
Mining and construction	581	6.1	2	8	27	48	10.3
Pharmaceutical	1404	1.9	0	0	3	41	15.5
Retail	650	0.5	0	0	3	8	1.8
Services	1141	3.0	0	1	11	90	12.6
Textiles and printing	319	12.4	0	1	66	280	46.1
Transportation	788	0.7	0	0	3	19	3.2
Others	272	1.3	0	2	6	11	2.5

The industry is classified based on a 4-digit SIC code. Agricultural, forestry, and fishing (SIC codes between 1 and 999); chemicals (SIC between 2800 and 2824, 2840 and 2899); computers (SIC between 7300 and 7379, 3570 and 3579, and 3670 and 3679); durable manufacturer (SIC between 3000 and 3999, excluding SIC between 3570 and 3579, 3670 and 3679); extractive industries (SIC between 2900 and 2999, 1300 and 1399); food (SIC between 2000 and 2111); mining and construction (SIC between 1000 and 1999, excluding SIC between 1300 and 1399); pharmaceutical (SIC between 2830 and 2836); retail (SIC between 5000 and 5999); services (SIC between 7000 and 8999, excluding SIC between 7370 and 7379); textiles and printing (SIC between 2200 and 2790); transportation (SIC between 4000 and 4899), and others. The patenting distribution is classified based on the quintile of PATENT, in which p50, p75, p95, and p99 represent quintile of 50%, 75%, 95%, and 99%, respectively

PATENTS are the real number of patents. BANKDEBT includes bank advances, bank liabilities, bank overdraft, current bank loans, notes payable to banks, and other borrowed funds reported as part of short-term borrowings, scaled by total assets. PUBLICDEBT includes corporate bonds (straight bonds, convertible bonds, and warrant bonds) divided by total assets. KEIRETSU_D is the dummy variable that equals one if the firm is affiliated with keiretsu. RD_TA is the R&D expenditure divided by total assets. TA is the real total assets (million dollars). LOGTA is the natural logarithm of total assets. AGE is the numbers of years since the firm was listed on the exchange. CAPEX_TA is the capital expenditure divided by total assets. TANGIBLE_TA is the ratio of tangibles assets to total assets. CASH_TA is cash and short-term investments divided by total assets. EBITDA_TA is the ratio of earnings before interests and taxes to total assets. NWC_TA is the ratio of working capital to total assets. O-SCORE = $-1.32 - 0.407 (\log \text{ total assets}) + 6.03 (\text{total liabilities/total assets}) - 1.43 (\text{working capital/total assets}) + 0.076 (\text{current liabilities/current assets}) - 1.72 (1 \text{ if total liabilities} > \text{total assets, } 0 \text{ otherwise}) - 0.521 ((\text{net income}_t - \text{net income}_{t-1}) / (|\text{net income}_t| + |\text{net income}_{t-1}|))$.

The average amount of total assets (TA) is 326,150 million USD, and the logarithm of total assets (LOGTA) is 10.99, which is higher than the values in 1994–2004 (4.3). These results indicate that Japanese capitalization has increased dramatically. Firm age (AGE) ranges from 1 to 65, and the average age is 34 years. The average RD_TA is 2.2%, which is lower than that in the USA (3.9%), but slightly higher than 1994–2004. The sample firms possess an average of 28% of tangible assets (including property, plants, and equipment). The average capital expenditure rate (CAPEX_TA) is 3.7%, which is higher than the 1.7% of the last 10 years. The cash and short-term investments ratio (CASH_R) indicates that the liquidity of Japanese firms is 16.7% on average. The mean of EBITDA_TA is 6.5%, revealing 6.5% of firms' financial profit over book assets. The sample firms possess an average of working capital over total assets (NWC_TA) that is 23.1%. The O_SCORE is the proxy to measure the firms' default probability. The average value of O_SCORE is -2.7 , indicating that the sample firms are financially sound.

Panel B of Table 1 reports the distribution of the number of patents granted among 13 industries based on the industry classifications. Specifically, our sample comprises agricultural, forestry, and fishing (SIC codes between 1 and 999); chemicals (SIC between 2800 and 2824, 2840 and 2899); computers (SIC between 7300 and 7379, 3570 and 3579, and 3670 and 3679); durable manufacturer (SIC between 3000 and 3999, excluding SIC between 3570 and 3579, 3670 and 3679); extractive industries (SIC between 2900 and 2999, 1300 and 1399), food (SIC between 2000 and 2111), mining and construction (SIC between 1000 and 1999, excluding SIC between 1300 and 1399); pharmaceutical (SIC between 2830 and 2836); retail (SIC between 5000 and 5999); services (SIC between 7000 and 8999, excluding SIC between 7370 and 7379); textiles and printing (SIC between 2200 and 2790); transportation (SIC between 4000 and 4899); and others. Computers firms show to be the most innovative industry among all, with an average patent of 56.8 for each firm, followed by durable manufacturing (24.6) and chemicals (10.4), in which 50% of firm-year observations have patents. Meanwhile, agricultural, forestry, and

fishing firms and retail firms are the least innovative as 95% of them do not have patents.

Figure 1 displays the average number of patents among listed firms over 2000–2015. Overall, the patent counts are increasing from 10.35 in 2000 to 33.19 in 2015, revealing a well-performed and growing innovative environment in our sample period. Figures 2 and 3 show a declining trend of bank debt and public debt and an ascending trend of R&D expenditure ratio over 2000–2015.

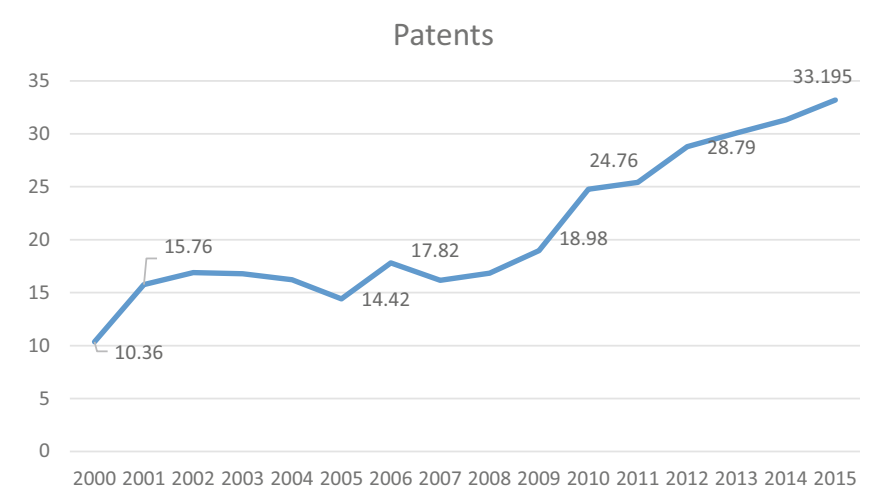


Fig. 1 Average number of patents among Japanese listed firms during 2000–2015. (Missing data is not included)

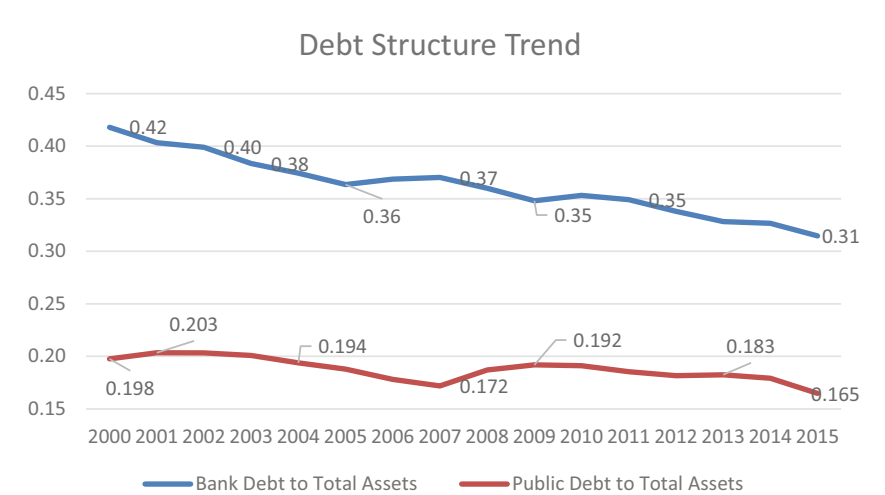


Fig. 2 Debt structure trend of Japanese firms during 2000–2015

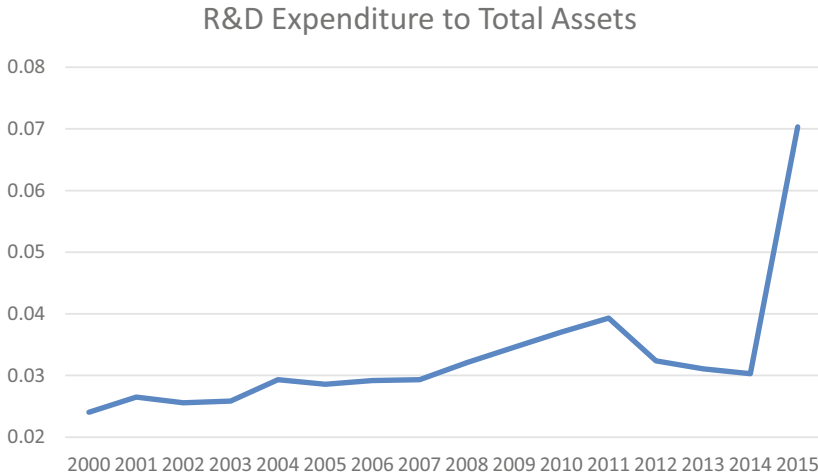


Fig. 3 R&D expenditure ratio of Japanese firms during 2000–2015

24.4.2 Univariate Tests

Panel A of Table 2 compares the characteristics between high- and low-innovative firms based on the average number of patents. The high-innovative firms show a higher ratio of keiretsu main bank relationships (22.8%) than low-innovative firms (18.4%), implying the support of the main banks. Besides, compared to low innovative firms, high innovative firms hold a significantly lower bank debt ratio (0.37 vs. 0.377) and a higher public debt ratio (0.156 vs. 0.14) and have relatively higher leverage (0.543 vs. 0.53). Moreover, high innovative firms are larger in size (11.39 vs. 10.54), are older (36 vs. 31), and have higher R&D investment (2.4% vs. 1.7%), capital expenditure ratio (3.9% vs. 3.4%), tangible assets ratio (0.297 vs. 0.281), working capital (0.219 vs. 0.214), and higher earnings (0.068 vs. 0.065). As high-innovative firms spend much more on R&D and physical assets, their cash and short-term investment are lower than low-innovative firms (0.16 vs. 0.17).

Panel A reports the descriptive statistics separately for high-innovative firms (PATENTS is larger than the Q3 of 1) and firms that are low innovative (PATENTS is lower than the Q3 of 1), and the differences in the means and their statistical significance based on a t-test. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. PATENTS are the number of patents granted. BANKDEBT includes bank advances, bank liabilities, bank overdraft, current bank loans, notes payable to banks, and other borrowed funds reported as part of short-term borrowings, scaled by total assets. PUBLICDEBT includes corporate bonds (straight bonds,

Table 2 Univariate tests

Panel A: High- and low-innovative firms							
Variables	Number of patents larger or equal to 1			Number of patents fewer than 1			Differences [1]–[4]
	(Obs. =8531)			(Obs. =32,768)			
	[1]	[2]	[3]	[4]	[5]	[6]	
	Mean	Median	Std. dev	Mean	Median	Std. dev	
PATENT	46.27	5	217.46	0	0	0	46.27***
KEIRETSU_D	0.215	0	0.256	0.073	0	0.23	0.142***
LEVERAGE	0.543	0.54	0.223	0.53	0.527	0.23	0.013***
BANKDEBT	0.370	0.402	0.173	0.377	0.356	0.202	−0.007*
PUBLICDEBT	0.156	134	0.103	0.14	0.126	0.12	0.026***
LOGTA	11.39	11.21	1.657	10.54	10.4	1.36	0.88***
AGE	36.91	40	19.02	31	35	20	5***
RD_TA	0.024	0.017	0.02	0.017	0.009	0.023	0.007***
CASH_TA	0.16	0.135	0.116	0.17	0.138	0.127	−0.013***
TANGIBLE_TA	0.297	0.285	0.149	0.281	0.271	0.157	0.016***
EBITDA_TA	0.068	0.055	0.042	0.065	0.05	0.053	0.003***
CAPEX_TA	0.039	0.032	0.03	0.034	0.024	0.034	0.005***
NWC_TA	0.219	0.217	0.183	0.214	0.211	0.196	0.005*
Panel B: Keiretsu and non-keiretsu firms							
Panel B reports the descriptive statistics for firms affiliated with the KEIRETSU main banks (KEIRETSU_D = 1) and firms that are not affiliated with KEIRETSU main banks (KEIRETSU_D = 0), and the differences in the means and their statistical significance based on a t-test. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively							
Variables	Keiretsu firms			Non-keiretsu firms			Differences [1]–[4]
	(Obs. =4221)			(Obs. =37,078)			
	[1]	[2]	[3]	[4]	[5]	[6]	
	Mean	Median	Std. dev	Mean	Median	Std. dev	
PATENT	58.37	1	233.1	4.002	0	42.39	54***
LEVERAGE	0.539	0.537	0.226	0.533	0.52	0.21	0.006*
BANKDEBT	0.362	0.36	0.188	0.355	0.35	0.196	0.007**
PUBLICDEBT	0.184	0.158	0.13	0.171	0.147	0.123	0.013***
LOGTA	11.49	11.29	1.72	10.89	10.72	1.501	0.6***
AGE	46.8	52	19.54	30	23	19.32	16.8***
RD_TA	0.022	0.014	0.025	0.02	0.013	0.024	0.002***
CASH_R	0.129	0.104	0.112	0.168	0.138	0.123	−0.039***
TANGIBLE_TA	0.291	0.279	0.157	0.29	0.281	0.151	0.001*
EBITDA_TA	0.063	0.054	0.048	0.06	0.055	0.047	0.003*
CAPEX_TA	0.038	0.03	0.031	0.037	0.028	0.033	0.001**
NWC_TA	0.223	0.195	0.183	0.197	0.222	0.19	0.026***

convertible bonds, and warrant bonds) divided by total assets. KEIRETSU_D is the dummy variable that equals 1 if the firm is affiliated with keiretsu. RD_TA is the R&D expenditure divided by total assets. TA is the real total assets (million USD). LOGTA is the natural logarithm of total assets. AGE is the numbers of years since the firm was listed on the exchange. CAPEX_TA is the capital expenditure divided by total assets. TANGIBLE_TA is the ratio of tangibles assets to total assets. CASH_TA is cash and short-term investments divided by total assets. EBITDA_TA is the ratio of earnings before interests and taxes to total assets. NWC_TA is the ratio of working capital to total assets.

Panel B of Table 2 compares the characteristics between keiretsu firms and non-keiretsu firms. The keiretsu firms show a higher number of patents than non-keiretsu firms (58 vs. 4), implying that keiretsu firms with group resources are more innovative. Besides, compared with non-keiretsu firms, keiretsu firms hold a significantly higher bank debt ratio (0.362 vs. 0.355) and public debt ratio (0.184 vs. 0.171) and have relatively higher leverage (0.539 vs. 0.533). Moreover, keiretsu firms are larger (11.49 vs. 10.89), older (46 vs. 30), and have slightly higher R&D investment (2.2% vs. 2.0%), higher capital expenditure ratio (3.8% vs. 3.7%), lower cash (0.129 vs. 0.168), higher tangible assets (0.291 vs. 0.29), higher earnings (0.063 vs. 0.06), and higher working capital (0.223 vs. 0.197).

24.4.3 2SLS with Instrumental Variable

Public debt use demonstrated to reveal the firms' credit quality in prior literature. Theory suggests that creditors are not always willing to finance investment with asymmetric information. The creditworthiness of a firm would mitigate the risk of asymmetric information. Myers and Majluf (1984) develop a framework to show that creditworthy firms are more likely to raise funds from bonds. Blackwell and Kidwell (1988) suggest that firms with high credit quality (creditworthiness) are more likely to borrow from the public than those with low credit quality. Denis and Minhov (2003) also find that creditworthy firms rely more on public debt financing than other debt. Since creditworthiness shows a firm's ability to pay back the debt to bondholders, it would affect public debt use. We employ credit quality, measured by O-score (Ohlson 1980), to instruct a firm's public debt use.²

Table 3 reports the 2SLS-IV results.

This table presents the testing for endogeneity problem of public debt use. We run the regression of public debt ratio (Eq. 1) on instrumental (O-SCORE) and all exogenous variables in column (1). In column (2), we estimate corporate innovation (Eq. 2) on debt structure, keiretsu affiliation, and the residual from Eq. (1). We then run the Durbin-Wu Hausman test to confirm the existence of endogeneity problem. We construct the instrumental O-score to mitigate the problem of endogeneity of

²The O-score is a predictor for a firm's bankruptcy probability. An O-score larger than 0.5 indicates a firm's high chance to default and low credit quality.

Table 3 Hausman test for endogeneity problem

	(1)	(2)
	2SLS – first stage	2SLS – second stage
Variable	PUBLICDEBT_HAT	INNOVATION
Instrument		
O-SCORE	−0.07*** (0.008)	
Debt structure		
BANKDEBT	−0.357*** (0.006)	−1.43*** (0.34)
PUBLICDEBT_HAT		4.59*** (0.756)
Keiretsu affiliation		
KEIRETSU_D	0.04*** (0.01)	1.46*** (0.41)
Firm characteristics		
LOGTA	0.012*** (0.0004)	0.649*** (0.025)
AGE	0.002*** (0.0007)	0.008* (0.024)
RD_TA	0.219*** (0.026)	16.7*** (0.886)
TANGIBLE_TA	0.106*** (0.006)	0.494*** (0.587)
EBITDA_TA	−0.22*** (0.014)	0.663 (0.464)
CASH_R	0.042*** (0.007)	−0.7*** (0.229)
CAPEX_TA	0.252*** (0.021)	0.632*** (0.673)
NWC_TA	0.02*** (0.007)	0.544*** (0.726)
Constant	−0.447*** (0.008)	5.2*** (0.779)
Industry effects	Yes	Yes
Year effects	Yes	Yes
Observations	26,353	26,353
R-squared	0.563	0.31
F-test for instrument (p-value)	0.000	
Durbin-Wu Hausman test (p-value)		0.000
p-value of F-test for the instrument is significant at the 1% level, indicating that the instrument is not weak		
p-value of Durbin-Wu Hausman test is significant at the 1% level, indicating that the endogeneity problem exists		

public debt ratio. $INNOVATION_{i,t}$ is the natural logarithm of the number of patents. $BANKDEBT_{i,t}$ includes bank advances, bank liabilities, bank overdraft, current bank loans, notes payable to banks, and other borrowed funds reported as part of short-term borrowings, scaled by total assets. $PUBLICDEBT_{i,t}$ includes corporate bonds (straight bonds, convertible bonds, and warrant bonds) divided by total assets. $KEIRETSU_D_{i,t}$ is the dummy variable that equals 1 if firm i is affiliated with keiretsu at year t . $RD_TA_{i,t}$ is the R&D expenditure divided by total assets. $LOGTA_{i,t}$ is the natural logarithm of total assets. $AGE_{i,t}$ is the natural logarithm of numbers of years since firm i was listed on the exchange. $CAPEX_TA_{i,t}$ is the capital expenditure divided by total assets. $TANGIBLE_TA_{i,t}$ is the ratio of tangibles assets to total assets. $CASH_R_{i,t}$ is cash and short-term investments divided by total assets. $EBITDA_TA_{i,t}$ is the ratio of earnings before interests and taxes to total assets. $O\text{-}SCORE_{i,t} = -1.32 - 0.407 (\log \text{ of total assets}) + 6.03 (\text{total liabilities/total assets}) - 1.43 (\text{working capital/total assets}) + 0.076 (\text{current liabilities/current assets}) - 1.72 (1 \text{ if total liabilities} > \text{total assets, } 0 \text{ otherwise}) - 0.521 ((\text{net income}_t - \text{net Income}_{t-1}) / (|\text{net income}_t| + |\text{net income}_{t-1}|))$. Standard errors appear in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Column (1) of Table 3 presents the first stage results with an instrument of O-score and control variables on the predicted public debt ($PUBLICDEBT_HAT$). The results indicate that O-score negatively and significantly affects the predicted public debt ($PUBLICDEBT_HAT$) at a significance level better than 1% (coefficient is -0.07). This result implies that low O-score firms borrow much more from public debt or that high credit quality firms tend to rely more on public debt borrowing. The F-test for the weakness of instrumental indicates a p-value of 0.001, revealing that our instrument is not weak. Column (2) of Table 3 shows the second stage results on corporate innovation by putting the predicted public debt ($PUBLICDEBT_HAT$). The Durbin-Wu Hausman test is significant, indicating an endogeneity problem of public debt exists. We, therefore, use 2SLS to solve this problem in the next section.

24.4.4 2SLS Regression Results

Table 4 reports the second-stage results of 2SLS. Results in Column (2) show that $PUBLICDEBT$ significantly fosters innovation with a coefficient of 4.36 ($p < 0.01$), indicating that a 1% increase in public debt ratio raises a firm's innovation by 4.36%. However, $BANKDEBT$ negatively reflects innovation at a significant level better than 1%; a 1% increase in bank debt ratio destroys the innovation by 1.4% (or 3 patents, where $3 = (e^{1.399} - 1)$). This finding suggests that bank financing provides smaller flexibility and risk tolerance for firms to adopt new techniques than public debt. We also find that although bank debt use hinders innovation, the keiretsu main bank relationship fosters innovation. The coefficient of $KEIRETSU_D$ is 1.57 ($p < 0.01$), implying that the keiretsu main bank relationships, on average, contributed 1.57% (or 3.8 patents) innovation to members. These findings are consistent with the univariate tests that a keiretsu firm possesses higher patents than non-keiretsu firms as the keiretsu main bank relationship mitigates asymmetric

Table 4 The 2SLS results of corporate innovation, debt structure, and keiretsu affiliation

	(1)	(2)
	2SLS – second stage	2SLS – second stage
Variables	INNOVATION	INNOVATION
Debt structure		
BANKDEBT	−1.401*** (0.271)	−1.399*** (0.316)
PUBLICDEBT	5.871*** (0.689)	4.365*** (0.732)
Keiretsu affiliation		
KEIRETSU_D		1.57*** (0.571)
Firm characteristics		
LOGTA	0.601*** (0.0147)	0.645*** (0.0144)
AGE	−0.000798 (0.00105)	0.000608 (0.000998)
RD_TA	15.44*** (0.824)	16.84*** (0.886)
EBITDA_TA	1.334*** (0.466)	0.556 (0.453)
NWC_TA	0.944*** (0.192)	0.475** (0.197)
CASH_R	9.973 (8.709)	10.38 (8.674)
CAPEX_TA	1.352*** (0.309)	0.614** (0.294)
TANGIBLE_TA	−1.584*** (0.597)	0.430 (0.589)
Industry effect	No	Yes
Year effect	No	Yes
Constant	−5.174*** (0.186)	−6.100*** (0.269)
Observations	26,358	26,358
R-squared	0.205	0.293

information and offers more sources for innovation projects (King and Levine 1993; De la Fuente and Marin 1996; Herrera and Minetti 2007). These results are also in line with Atanassov's (2016) reports showing that public debt spurs innovation. Figures 4 and 5 depicted the scatter diagram of public debt, bank debt ratio, and corporate innovation with fitted lines, respectively. The results of fitted lines are under our expectations.

This table presents the second-stage results of 2SLS regressions, in which the corporate innovation is estimated by the equation:

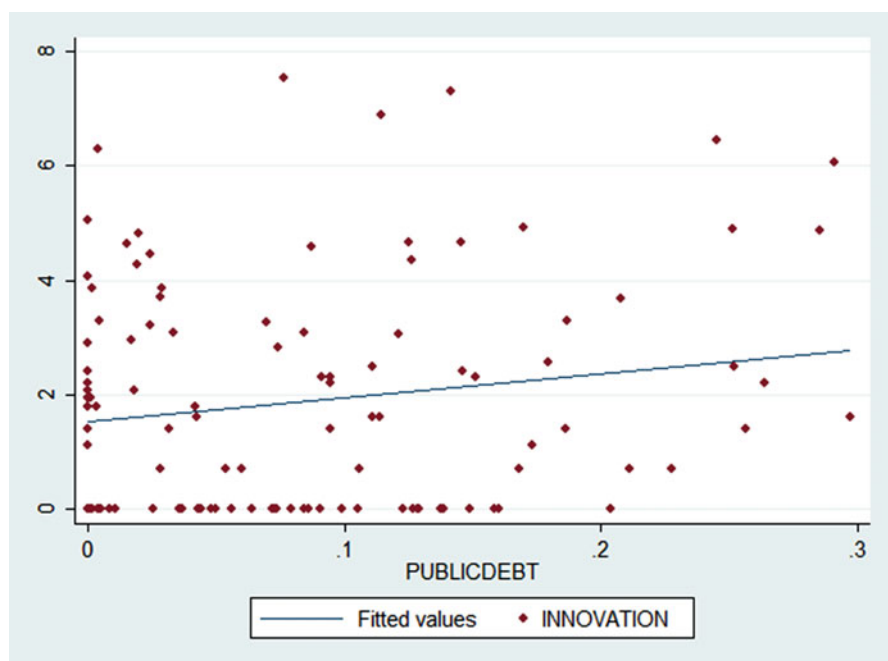


Fig. 4 Relationship between public debt ratio and corporate innovation

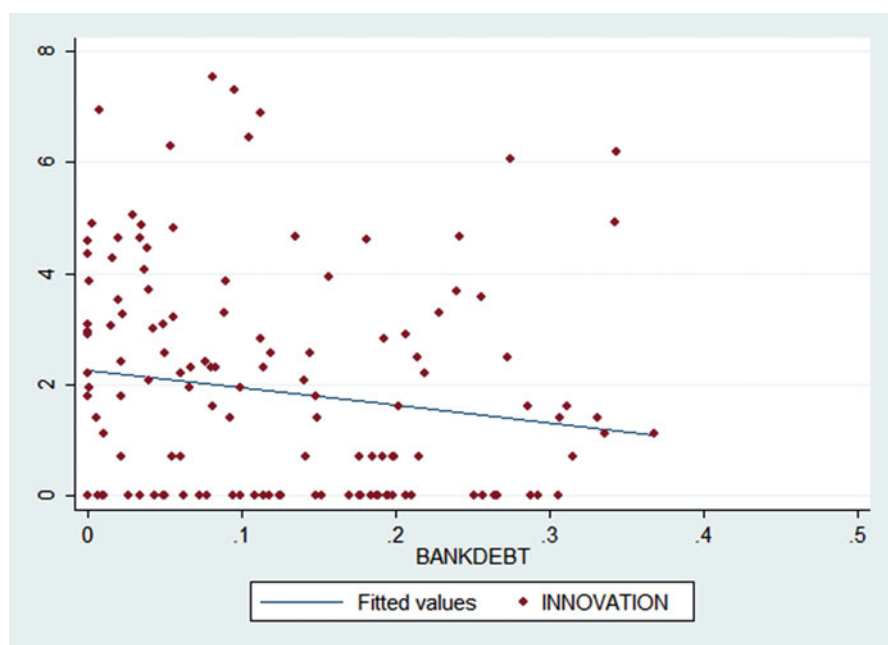


Fig. 5 Relationship between bank debt ratio and corporate innovation

$$\begin{aligned}
INNOVATION_{i,t} = & \beta_0 + \beta_1 BANKDEBT_{i,t} + \beta_2 PUBLICDEBT_{i,t} \\
& + \beta_3 KEIRETSU_D_{i,t} + \alpha Controls_{i,t} + Industry_i + Year_t \\
& + \varepsilon_{i,t}
\end{aligned} \tag{3}$$

We use O-score as an instrument to instruct the public debt to mitigate the endogeneity of debt structure in the first stage. $INNOVATION_{i,t}$ is the natural logarithm of the number of patents. $BANKDEBT_{i,t}$ includes bank advances, bank liabilities, bank overdraft, current bank loans, notes payable to banks, and other borrowed funds reported as part of short-term borrowings, scaled by total assets. $PUBLICDEBT_{i,t}$ includes corporate bonds (straight bonds, convertible bonds, and warrant bonds) divided by total assets. $KEIRETSU_D_{i,t}$ is the dummy variable that equals 1 if firm i is affiliated with keiretsu at year t . $RD_TA_{i,t}$ is the R&D expenditure divided by total assets. $LOGTA_{i,t}$ is the natural logarithm of total assets. $AGE_{i,t}$ is the natural logarithm of numbers of years since firm i was listed on the exchange. $CAPEX_TA_{i,t}$ is the capital expenditure divided by total assets. $TANGIBLE_TA_{i,t}$ is the ratio of tangibles assets to total assets. $CASH_R_{i,t}$ is cash and short-term investments divided by total assets. $EBITDA_TA_{i,t}$ is the ratio of earnings before interests and taxes to total assets. In parentheses are standard errors based on robust standard errors adjusted for heteroscedasticity and firm-level clustering. The notations of *, **, and *** represent statistical significance at the 10%, 5%, and 1% levels, respectively.

In terms of the control variables, we find that firm size, working capital, R&D, and capital expenditure are positively related to corporate innovation. The coefficient of RD_TA is 15.22, indicating that a 1% increase in R&D expenditure leads to a rise in innovation by 15.22%. Large firms and high profitability firms are likely to be more innovative as these firms appear to have more capital to invest in innovation projects than their counterparts. $CAPEX_TA$ and $TANGIBLE_TA$ are also positively related to innovation, revealing that firms that spend more on physical assets and have more tangible assets would create more patents. Firms with high NWC_TA have more patents as sufficient short-term investments than liquidity debts and are financially sound. The cash ratio also positively affects innovation but insignificantly; the plausible reason may relate to our sample.³ Industries and years factors also affect the firms' innovation.

24.5 Robustness Checks

24.5.1 The Exogenous Shocks of Keiretsu Consolidation on Corporate Innovation

We have demonstrated that keiretsu contributes to more patents for Japanese firms. To clarify the concern about the possibility that the consolidated keiretsu might drive corporate innovation, we construct a DID (difference-in-differences) analysis of

³ As half of our samples have no patents and these firms have high cash ratio, these observations may affect the result.

keiretsu consolidation on corporate innovation in this section to explore whether more patents are related to keiretsu consolidation exogenous shock.

During the 1990s, Japan suffered from the “asset price bubble,” in which real estate and stock market prices were substantially inflated during 1986–1991 and then collapsed in early 1992. The recession was caused by the excessive loan growth quotas, in which Japanese banks lent heavily to the real estate sector and even held land directly. When the real estate value dropped, they incurred significant losses on their loans and their lands. Many banks, including the large ones, were hit hard by bad loan portfolios and forced to merge or go out of business. In 2000, Sumitomo Bank was merged with Sakura Bank, part of the Mitsui group. In 2001, Sanwa Bank (Sanwa group) merged with Tokai Bank (Tokai group) to form UFJ Bank. In 2002, Fuji Bank (Fuyo group) joined with Dai-Ichi Kangyo Bank and the independent bank – Industrial Bank of Japan – to create the Mizuho group. In 2006, Bank of Tokyo-Mitsubishi and UFJ Bank were merged to form the Bank of Tokyo-Mitsubishi UFJ. The consolidation led to tie-ups, large entities, and boosted scale of business in the traditional keiretsu. Before the merger, many banks were undercapitalized; the merger improved their financial strength and their active corporate governance roles in monitoring affiliated firms.

Our DID model to estimate the effect of consolidated keiretsu on innovation is below:

$$\begin{aligned} INNOVATION_{i,t} = & \beta_0 + \beta_1 BANKDEBT_{i,t} + \beta_2 PUBLICDEBT_{i,t} \\ & + \beta_3 KEIRETSU_D_{i,t} + \beta_4 KEIRETSU_M_{i,t} * POST_M \\ & + \alpha Controls_{i,t} + Industry_i + \varepsilon_{i,t} \end{aligned} \quad (4)$$

where $KEIRETSU_M$ is a dummy that equals 1 for firms affiliated to the consolidated keiretsu, in which two banks from separate keiretsu merged, and 0 otherwise. For example, NSK Ltd. (originally Fuyo group) indicates a $KEIRETSU_M$ that equals 1 because of the merging of Fuji Bank (Fuyo group) and Dai-Ichi Kangyo Bank (Dai-Ichi Kangyo group) and Industry Bank of Japan. $POST_M$ is a dummy that equals 1 for observations of the consolidated keiretsu after bank merging and 0 otherwise. The interaction term $KEIRETSU_M * POST_M$ captures the magnitude of the effect of the consolidated keiretsu after the merger, compared with those that did not. Other control variables, time-varying, and industry effects are similar to the main regression. Table 5 presents the results.

This table presents the second-stage results of 2SLS regressions of corporate innovation while considering the exogenous shock of the event of keiretsu mergers, in which the corporate innovation is estimated by the equation Eqs. (1) and (4). $KEIRETSU_M_{i,t}$ is the dummy variable, equals 1 for keiretsu firms experiencing consolidation and zero otherwise. $POST_M_{i,t}$ equals 1 for the year after the keiretsu merging and 0 otherwise. In parentheses are standard errors based on robust standard errors adjusted for heteroscedasticity and firm-level clustering. *, **, and *** represent statistical significance at the 10%, 5%, and 1% levels, respectively.

Column (1) of Table 5 reports the DID effect of consolidated keiretsu on the natural logarithm of patents in the model. Columns (2)–(4) repeat the baseline model by adding debt structure and keiretsu affiliation, firm characteristics, and time-

Table 5 2SLS regression of corporate innovation on debt structure, keiretsu affiliation, and keiretsu consolidation

	(2)	(3)	(4)
	2SLS – second stage	2SLS second stage	2SLS – second stage
Variables	INNOVATION	INNOVATION	INNOVATION
Debt structure:			
BANKDEBT	−9.857***	−1.4***	−1.38***
	(0.413)	(0.276)	(0.324)
PUBLICDEBT	29.829***	5.895***	4.37***
	(1.08)	(0.696)	(0.731)
Keiretsu consolidation:			
KEIRETSU_M*POST_M	3.76***	2.98***	2.87***
	(1.44)	(1.01)	(0.95)
KEIRETSU_M	1.58	−2.48***	−3.09***
	(1.44)	(0.95)	(0.91)
Firm characteristics:			
LOGTA		0.235**	0.647***
		(0.009)	(0.0144)
AGE		−0.004***	0.0006
		(0.001)	(0.001)
RD_TA		15.47***	16.90***
		(0.825)	(0.886)
EBITDA_TA		1.361***	0.564
		(0.467)	(0.453)
NWC_TA		0.924***	0.463**
		(0.192)	(0.197)
CASH_R		−3.519***	−1.738***
		(0.570)	(0.544)
CAPEX_TA		1.319***	0.572*
		(0.309)	(0.294)
TANGIBLE_TA		−1.512**	0.469
		(0.598)	(0.589)
Industry effect	No	No	Yes
Year effect	No	No	Yes
Constant	1.319***	−5.161***	−6.122***
	(0.0742)	(0.187)	(0.270)
Observations	26,358	26,358	26,358
R-squared	0.187	0.206	0.295

varying and industry-fixed effects, respectively. Our main focus is the coefficient of the interaction term KEIRETSU_M*POST_M, which captures the DID impact on the changes in innovation over time between firms experiencing a keiretsu consolidation and those that did not. As columns (3)–(4) show, the direct effects of KEIRETSU_M are all significantly negative, with coefficients of −2.48

($p < 0.01$) and -3.09 ($p < 0.01$), implying that the endogeneity concerns about the direct effect of consolidation event drive the results does not exist.

The interactive effect of $\text{KEIRETSU_M} \times \text{POST_M}$ after the merger becomes significantly positive at the 1% level in columns (1)–(4), implying that the consolidated keiretsu has increased innovation among its members after the merger. The coefficient of $\text{KEIRETSU_M} \times \text{POST_M}$ on innovation is 2.87 in column (4) and also implies that the consolidated keiretsu increased innovation among its members by 2.87% compared to their counterparts as more resource ties to the firms. The effects of public debt and bank debt on innovation remain significant without change compared to the baseline results. These results clarify the concerns from exogenous shocks and are in line with the previous conclusions.

24.5.2 Propensity Score Matching

We conduct propensity score matching proposed by Rosenbaum and Rubin (1983) to mitigate the self-selection bias concerns based on the unobserved characteristics. The matching pair of keiretsu and non-keiretsu firms are selected based on the similarities in observed characteristics.

We begin the score-matching analysis based on the debt structure of firms. In Panel A of Table 6, we match keiretsu and non-keiretsu firms with bank debt ratio and public debt ratio and computed the differences in innovation. Keiretsu firms indicate an average innovation of 2.476, which is greater than that of non-keiretsu firms (2.025). The difference is highly significant. In Panel B, we generate propensity-score-matching firms based on firm characteristics (e.g., total assets, tangibility assets, capital expenditure, cash and short-term investments, earnings before interests and rate, working capital, etc.). The matched keiretsu and non-keiretsu pair observations are 1454, respectively. The results are consistent with Panel A, in which firms that affiliated with keiretsu main banks show significantly larger patents than non-keiretsu firms. Panel C matches keiretsu affiliation with debt structure, firm characteristics, industry, and time factors. The findings show a significantly higher innovation of firms affiliated with keiretsu over non-affiliated firms at a 1% significance level. Our matching approach uses the Logit

Table 6 Propensity score matching

	With keiretsu	With no keiretsu	Difference
<i>Panel A: Matched by debt structure (n = 1454)</i>			
INNOVATION	2.476	2.025	0.451***
<i>Panel B: Matched by firm characteristics (n = 1454)</i>			
INNOVATION	2.47	2.19	0.17***
<i>Panel C: Matched by debt structure and firm characteristics, industry controls, and time controls (n = 1361)</i>			
INNOVATION	2.526	2.01	0.516***

model with a likelihood ratio test for bias. The results show that keiretsu firms' innovation is significantly higher than non-keiretsu firms.

Table 5 measures the effect of keiretsu affiliation on corporate innovation after controlling for debt structure and firm characteristics. We match each firm with keiretsu main bank relationship to a firm with no keiretsu relationship. Panel A requires matches to have the same bank debt and public debt ratio. Panel B is similar to the matching procedure in Panel A but adds firm characteristics (natural log total assets, capital expenditure, assets tangibility, working capital, profitability, cash, and short-term investments). Panel C adding industry controls and time control to Panel B. INNOVATION defined as the natural logarithm of patents. BANKDEBT measured by current liabilities divided by total assets. BANKDEBT includes bank advances, bank liabilities, bank overdraft, current bank loans, notes payable to banks, and other borrowed funds reported as part of short-term borrowings, scaled by total assets. PUBLICDEBT includes corporate bonds (straight bonds, convertible bonds, and warrant bonds) divided by total assets. Firm characteristics include firm scale (total assets, unit: million dollars), the capital expenditure divided by total assets (CAPEX_TA), the ratio of tangibles assets to total assets (TANGIBLE_TA), cash and short-term investments divided by total assets (CASH_R), the ratio of earnings before interests and taxes to total assets (EBITDA_TA), and the ratio of working capital to total assets (NWC_TA). The significance level of differences based on t-test. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

24.5.3 Other Measurements of Corporate Innovation

We use an alternative innovation measure, namely, innovation efficiency, to reconfirm the robustness check on innovation. Innovation efficiency aims at both innovation input and output, reflecting whether the firm creates these patents at a reasonable cost. The innovation efficiency is measured by the productivity of the output (patents number) relative to input (R&D) (see Hirshleifer et al. 2013). This measurement allows investors to evaluate a firm's efficiency of R&D in the form of patents granted per dollar of R&D expenditure. The $IE_{i,t}$ is defined as Eq. (5):

$$IE_{i,t} = \frac{(\text{Patents}_{i,t})}{(R\&D_{i,t-1} + 0.8 * R\&D_{i,t-2} + 0.6 * R\&D_{i,t-3} + 0.4 * R\&D_{i,t-4} + 0.2 * R\&D_{i,t-5})} \quad (5)$$

where the innovation efficiency at time t is measured by the patents at time t divided by the cumulative R&D capital from years t-1 to t-5 with a 20% depreciation rate per year (see Chan et al. 2001; Lev et al. 2005) denotes the amounts of R&D investments of firm i in year t-1 and so on. The revised baseline equation is specified as (6) below:

$$\begin{aligned} LOG_IE_{i,t} = & \beta_0 + \beta_1 BANKDEBT_{i,t} + \beta_2 PUBLICDEBT_{i,t} \\ & + \beta_3 KEIRETSU_D_{i,t} + \alpha Controls_{i,t} + Industry_i + Year_t + \varepsilon_{i,t} \end{aligned} \quad (6)$$

Table 7 reports the innovation efficiency results. $\text{LOG_IE}_{i,t}$ is the natural logarithm of $\text{IE}_{i,t}$. Consistent with the prior results, BANKDEBT is significantly and negatively related to firm innovation efficiency in columns (1) and (2), further confirming that borrowing loans from banks hinders innovation. Similarly, PUBLICDEBT is significantly and positively related to corporate innovation at the 1% level, reconfirming that firms that use more public debt have higher innovation efficiency. The coefficient of KEIRETSU_D is positive and significant at 0.41, reconfirming the keiretsu main bank relationship significantly raised innovation efficiency among its members by 0.41.

Table 7 2SLS results of innovation efficiency, debt structure, and keiretsu affiliation

	(1)	(2)
	2SLS – second stage	2SLS – second stage
Variables	LOG_IE	LOG_IE
Debt structure		
BANKDEBT	−0.047 (0.043)	−0.135** (0.0530)
PUBLICDEBT	0.48*** (0.109)	0.562*** (0.123)
Keiretsu affiliation		
KEIRETSU_D	0.32*** (0.09)	0.41*** (0.101)
Firm characteristics		
LOGTA	−0.012*** (0.002)	−0.014*** (0.00242)
AGE	−0.0005*** (0.0001)	−0.0004** (0.000168)
RD_TA	0.041 (0.13)	0.314** (0.149)
EBITDA_TA	0.191*** (0.073)	0.234*** (0.0760)
NWC_TA	0.037 (0.032)	−0.0297 (0.0330)
CASH_R	0.167* (0.09)	0.227** (0.0912)
CAPEX_TA	−0.069 (0.05)	−0.0826* (0.0493)
TANGIBLE_TA	0.011 (0.094)	0.111 (0.0988)
Industry effect	No	Yes
Year effect	No	Yes
Constant	0.115*** (0.029)	0.0787* (0.0452)
Observations	26,358	26,358

This table reports the 2SLS regression results of corporate innovation efficiency on debt structure and keiretsu affiliation.

$$\begin{aligned} \text{LOG_IE}_{i,t} = & \beta_0 + \beta_1 \text{BANKDEBT}_{i,t} + \beta_2 \text{PUBLICDEBT}_{i,t} \\ & + \beta_3 \text{KEIRETSU_D}_{i,t} + \beta_4 \text{Controls}_{i,t} + \text{Industry}_i + \text{Year}_t + \varepsilon_{i,t} \quad (7) \end{aligned}$$

The dependent variable is applied to an alternative measure of corporate innovation. Instead of using patents, innovation efficiency, which accounts for both innovation input and output, is used to proxy for corporate innovation. The endogeneity problem of public debt ratio (PUBLICDEBT) is mitigated by instrumental Ohlson O-score. In parentheses are standard errors based on robust standard errors adjusted for heteroscedasticity and firm-level clustering. The notations of *, **, and *** represent statistical significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is LOG_IE_{i,t}, defined as the natural logarithm of innovation efficiency for firm *i* in year *t*. All the variables are measured by firm-year. BANKDEBT_{i,t} includes bank advances, bank liabilities, bank overdraft, current bank loans, notes payable to banks, and other borrowed funds reported as part of short-term borrowings, scaled by total assets. PUBLICDEBT_{i,t} includes corporate bonds (straight bonds, convertible bonds, and warrant bonds) divided by total assets. KEIRETSU_D_{i,t} is the dummy variable that equals one if firm *i* is affiliated with keiretsu at year *t*. RD_TA_{i,t} is the R&D expenditure divided by total assets. LOGTA_{i,t} is the natural logarithm of total assets. AGE_{i,t} is the natural logarithm of numbers of years since firm *i* was listed on the exchange. CAPEX_TA_{i,t} is the capital expenditure divided by total assets. TANGIBLE_TA_{i,t} is the ratio of tangibles assets to total assets. CASH_R_{i,t} is cash and short-term investments divided by total assets. EBITDA_TA_{i,t} is the ratio of earnings before interests and taxes to total assets.

24.6 Conclusions

Understanding the effect of debt on innovation activities is of great importance for firm sustainability and long-term economic growth. Firms usually access external finance through debts to facilitate innovation and corporate growth. However, the debt financing mixture may diminish corporate innovation if firms do not possess an appropriate debt structure. We find that public debt use fosters corporate innovation because of more risk tolerance and financial flexibility in financing from bondholders. The use of bank debt, by contrast, hinders innovation. This result may relate to the diligent monitoring from banks associated with the loan covenants that provide less flexibility for firms to beat the potential benefits with upside uncertainty. We find that once the firms have keiretsu main bank relationships, it provides more risk tolerance. This tolerance associated with the ownerships held by banks switches the keiretsu main bank to act more like shareholders than creditors, spurring more innovation. The postconsolidation keiretsu banks become more supportive of their member firms and raise the innovation by 2.87%.

Our conclusions provide broad implications for managers, banks, and policy-makers while raising the external funds for innovative projects.

References

- Acharya, V., and Z. Xu. 2017. Financial dependence and innovation: The case of public versus private firms. *Journal of Financial Economics* 124 (2): 223–243.
- Adam, T., and D. Streitz. 2016. Hold-up and the use of performance-sensitive debt. *Journal of Financial Intermediation* 26: 47–67.
- Aflonso, E., and S. Giannangeli. 2012. *Outsourcing innovation and the role of bank debt for SMEs*. Uni Credit Best Paper Award.
- Aghion, P., and P. Howitt. 1992. A model of growth through creative destruction. *Econometrica* 60: 323–351.
- Aghioon, P., and J. Tilore. 1994. The Management of Innovation. *The Quarterly Journal of Economics* 109 (4): 1185–1209.
- Atanassov, J. 2016. Arm's length financing and innovation: Evidence from publicly trade firms. *Management Science* 62 (1): 128–155.
- Backman, M., and T. Wallin. 2018. Access to banks and external capital acquisition: Perceived innovation obstacles. *Annals of Regional Science* 61 (1): 161–187.
- Brancati, E. 2015. Innovation financing and the role of relationship lending for SMEs. *Small Business Economics* 44 (2): 449–473.
- Brown, J.R., S.M. Fazzari, and B.C. Petersen. 2009. Financing innovation and growth: Cash flow, external equity, and the 1990s R&D boom. *Journal of Finance* 64 (1): 151–185.
- Brown, J.R., G. Martinsson, and B.C. Petersen. 2013. Law, stock markets, and innovation. *Journal of Finance* 68: 1517–1549.
- Campbell, Tim S., and A.K. William. 1980. Information production, market signaling, and the theory of financial intermediation. *Journal of Finance* 35: 863–882.
- Chan, L.K., J. Lakonishok, and T. Sougiannis. 2001. The stock market valuation of research and development expenditures. *The Journal of Finance* 56 (6): 2431–2456.
- Chava, S., X. Chong, and V. Nanda. 2012. Funding innovation: The role of lender expertise and control rights, Georgia Institute of Technology. *NBER Working Paper* 20676.
- Cornaggia, J., Y. Mao, X. Tian, and B. Wolfe. 2015. Does banking competition affect innovation? *Journal of Financial Economics* 115 (1): 189–209.
- Czarnitzki, D., and K. Kraft. 2007. Are credit ratings valuable information? *Applied Financial Economics* 17 (13): 1061–1070.
- De la Fuente, A., and J.M. Marin. 1996. Innovation, bank monitoring and endogenous financial development. *Journal of Monetary Economics* 38 (2): 269–301.
- De Mel, S., D. McKenzie, and C. Woodruff. 2009. Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of Development Economics* 88 (1): 19–31.
- Denis, D.J. and V. Minhov. 2003. The choice among bank debt, non-bank private debt and public debt: Evidence from new corporate borrowings. *Journal of Financial Economics* 70 (1): 3–28.
- Diamond, D. 1984. Financial intermediation and delegated monitoring. *Review of Economic Studies* 51: 393–414.
- . 1991. Monitoring and reputation: The choice between bank loans and directly placed debt. *Journal of Political Economy* 99: 689–721.
- Fama, E. 1985. What's different about banks? *Journal of Monetary Economics* 15: 29–39.
- Gu, Y.Q., C.N.X. Mao, and X. Tian. 2017. Banks' interventions and firms' innovation: Evidence from debt covenant violations. *Journal of Law & Economics* 60 (4): 637–671.
- Hasan, M.B., A.F.M.M. Ahsan, M.A. Rahaman, and M.A. Nurul. 2014. Influence of capital structure on firm performance: Evidence from Bangladesh. *International Journal of Business and Management* 9 (5): 184–194.
- Herrera, A.M., and R. Minetti. 2007. Informed finance and technological change: Evidence from credit relationships. *Journal of Financial Economics* 83: 223–269.

- Hirshleifer, D., P.-H. Hsu, and D. Li 2013. Innovative efficiency and stock returns. *Journal of Financial Economics* 107 (3): 632–654.
- Holmstrom, B. 1989. Agency costs and innovation. *Journal of Economic Behavior and Organization* 12: 305–327.
- Hoshi, T., and A. Kashyap. 2001. *Corporate financing and governance in Japan* Cambridge, Massachusetts: The MIT Press.
- Hoshi, T., A.K. Kashyap, and D.S. Scharfstein. 1993. The choice between public and private debt: An analysis of post-deregulation corporate financing in Japan. NBER Working Paper No. w4421.
- Jensen, M.C. 1986. Agency cost of free cash flow, corporate finance, and takeovers. *American Economic Review* 76 (2).
- Kerr, W.R., and R. Nanda. 2015. Financing innovation. *Annual Review of Financial Economics* 7 (1): 445–462.
- Kim Minjung and Park Jungsoo 2017. Do bank loans to financially distressed firms lead to innovation? *The Japanese Economic Review*, Springer 68 (2): 244–256.
- King, R.G., and R. Levine. 1993. Finance, entrepreneurship, and growth: Theory and evidence. *Journal of Monetary Economics* 32 (3): 513–542.
- Leland, H.E., and D. Pyle. 1977. Informational asymmetries, financial structure and financial intermediation. *Journal of Finance* 32: 371–387.
- Lev, B., B. Sarath, and T. Sougiannis 2005. R&D reporting biases and their consequences. *Contemporary Accounting Research* 22 (4): 977–1026.
- Lim, S., S.C. Mann, and V.T. Mihov. 2017. Do operating leases expand credit capacity? Evidence from borrowing costs and credit ratings. *Journal of Corporate Finance* 42: 100–114.
- Mancusi, M.L., A. Vezzulli, S. Frazzoni, Z. Rotondi, and M. Sobrero. 2018. Export and innovation in small and medium enterprises: The role of concentrated bank borrowing. *Economica* 85 (337): 177–204.
- Mann, W. 2018. Creditor rights and innovation: Evidence from patents collateral. *Journal of Financial Economics* 130 (1): 25–47.
- Manso, G. 2011. Motivating innovation. *Journal of Finance* 66: 1823–1869.
- Mattes, J.A., S. Steffen, and M. Wahrenburg. 2013. Do information rents in loan spreads persist over the business cycles? *Journal of Financial Services Research* 43 (2): 175–195.
- Myers, S.C., and N. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13: 187–221.
- Nanda, R., and T. Nicholas. 2014. Did bank distress stifle innovation during the great depression? *Journal of Financial Economics* 114 (2): 273–292.
- Ogawa, K. 2007. Debt, R&D investment and technological progress: A panel study of Japanese manufacturing firms' behavior during the 1990s. *Journal of Japan International Economics* 21 (4): 403–423.
- Ohlson, J.A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18: 109–131.
- Rajan, R. 1992. Insiders and outsiders: the choice between informed and arm's length debt. *Journal of Finance* 47: 1367–1400.
- Rajan, R., and L. Zingales. 2003. The great reversals: The politics of financial development in the twentieth century. *Journal of Financial Economics* 69: 5–50.
- Robb, A.M., and D.T. Robinson. 2014. The capital structure decisions of new firms. *Review of Financial Studies* 27 (1): 153–179.
- Rosenbaum, P.R., and D.B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1): 41–55.
- Saidi, F., and A. Zaldokas. 2021. How does firms' innovation disclosure affect their banking relationships? *Management Science* 67 (2): 742–768.
- Santos, J.A.C., and A. Winton. 2008. Bank loans, bonds, and information monopolies across the business cycle. *Journal of Finance* 63 (3): 1315–1359.
- Sharpe, S.A. 1990. Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *Journal of Finance* 45: 1069–1087.
- Sumit Agarwal, S., and R. Hauswald. 2010. Distance and private information in lending. *Review of Financial Studies* 23 (7): 2757–2788.

- Tadesse, S. 2006. Innovation, information and financial architecture. *Journal of Financial and Quantitative Analysis* 41 (4): 753–786.
- Tian, X., and T.Y. Wang. 2014. Tolerance for failure and corporate innovation. *Review of Financial Studies* 27: 211–255.
- Titman, S., and R. Wessels. 1988. The determinants of capital structure choice. *Journal of Finance* 43: 1–19.
- Weinstein, D., and Y. Yafeh. 1998. On the costs of a bank-centered financial system: Evidence from the changing main bank relations in Japan. *Journal of Finance* 53: 635–672.
- Yu, H.C., C.S. Chen, and D.Z. Hsieh. 2010. Keiretsu style main bank relationships, r&d investment, leverage, and firm value: Quantile regression approach. In *Handbook of quantitative finance and risk management*, 829–841. Springer.
- Zheng, G.P., S.X. Wang, and Y.X. Xu. 2018. Monetary stimulation, bank relationship and innovation: Evidence from China. *Journal of Banking & Finance* 89: 237–248.



Thomas S. Y. Ho and Sang Bin Lee

Abstract

Interest movement models are important to financial modeling because they can be used for valuing any financial instruments whose values are affected by interest rate movements. Specifically, we can classify the interest rate movement models into two categories: equilibrium models and no-arbitrage models. The equilibrium models emphasize the equilibrium concept. However, the no-arbitrage models argue that the term-structure movements should satisfy the no-arbitrage condition. The arbitrage-free interest rate model is an extension of the Black–Scholes model to value interest rate derivatives. The model valuation is assured to be consistent with the observed yield curve in valuing interest rate derivatives and providing accurate pricing of interest rate contingent claims. Therefore, it is widely used for portfolio management and other capital market activities.

Keywords

Black, Derman, and Toy model · Brennan and Schwartz two-factor model · Cox, Ingersoll and Ross model · Ho and Lee model · Hull and White model · Interest correlation · Lognormal versus normal movements · Mean reversion · Term structure volatility · Vasicek model

This chapter is from *The Oxford Guide to Financial Modeling: Applications for Capital Markets, Corporate Finance, Risk Management, and Financial Institutions* by Thomas Ho and Sang Bin Lee, copyright © 2004 by Thomas S.Y. Ho and Sang Bin Lee. Used with permission of Oxford University Press, Inc.

T. S. Y. Ho (✉)

Thomas Ho Company, Ltd., New York, NY, USA

e-mail: Tom.Ho@ThomasHo.com

S. B. Lee

Hanyang University, Seoul, South Korea

25.1 Introduction

There are many examples of interest rate derivatives that are actively traded in over-the-counter markets and in organized exchanges. Caps, floors, Treasury bond options, Treasury bond futures options, Euro-dollar futures options, and swaption are just some examples of this important class of derivatives in our financial markets. They are classified as “interest rate derivatives” because their stochastic movements are directly related to the interest rate movements in a way that is analogous to the stock option price that moves in step with the underlying stock price.

We first present an empirical analysis of historical yield curve movements, which conveys its relationship to interest rate models. Then we provide an overview of the interest rate models.

25.2 Interest Rate Movements: Historical Experiences

Interest rate movements refer to the uncertain movements of the Treasury spot yield curve. Each STRIPS bond is considered a security. When the daily closing price is reported, the bond’s yield-to-maturity can be calculated. The observed Treasury spot yield curve is the scattered plot of the yield to maturity against the maturity for all the STRIPS bonds. Since the spot yield curve is a representation of the time value of money, and the time value of money is related to the time-to-horizon in a continuous fashion, the scattered plots should be a continuous curve. Hence, we call the scattered plot a yield curve.

What are the dynamics of the spot yield curve? Let us consider the behavior of spot yield curve movements in relation to interest rate levels, historically. The monthly spot yield curves from the beginning of 1994 until the end of 2001 are depicted in the figure below.

As Fig. 1 shows, the spot yield curves can take on a number of shapes. When the yields of the bonds increase with the bonds’ maturities, the yield curve is said to be upward sloping. Conversely, when the yield decreases with maturity, the spot curve is called downward sloping. Although not shown in Fig. 1, the early 1980s displayed a yield curve that was downward sloping. In 1998, the yield curve was level or flat. In the early part of 2001, the yield curve was humped, with the yields reaching the peak at the 1 year maturity. Historically, the spot yield curve has changed its shape as well as the level continually.

The yield curve movement is concerned with the change of the yield curve shape over a relatively short time interval, say, 1 month. Describing yield curve movements is slightly more complicated than describing a stock movement. To describe the movement of stocks, we can decompose the stock movement into two parts: the expected drift or expected returns and the uncertain movement. The model is represented by:

$$dS = \mu S dt + \sigma S dZ \quad (1)$$

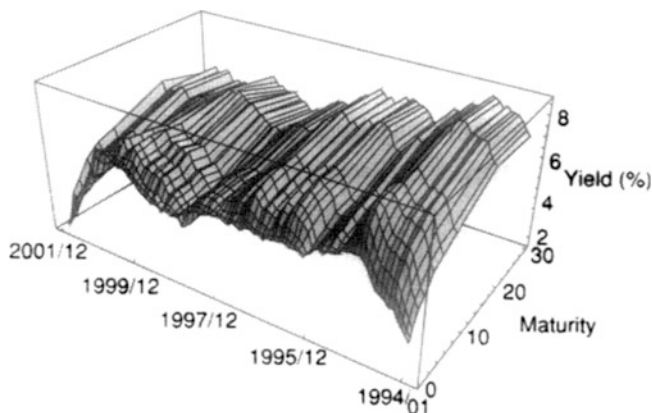


Fig. 1 A time-series diagram of monthly spot yield curve movements (1994.01–2001.12). (Data source: <http://www.economagic.com/>)

where dS represents a small movement for a short time interval dt . μ is called the instantaneous returns of the stock, σ is the instantaneous standard deviation (or volatility) of the stock. dZ represents a small uncertain movement specified by a normal distribution. The mean and the standard deviation of the normal distribution is 0 and $\sqrt{dt}$, respectively. The first term is called the drift term. It represents the expected movement of the stock price. If the first term is zero, then the future stock price is *expected* to remain the same as the present observed price. Of course, the realized stock price in the future can deviate from the initial stock price because of the uncertain stock price movement specified by the second term. The random term dZ can be viewed as a unit of risk, a normal distribution over an (infinitely) short time interval. The coefficient of the dZ term represents the volatility of the process. If this coefficient is zero, then the process has no risk, and the stock price movement has no uncertainty.

But to specify the movement of the yield curve, in a way that is similar to Eq. (1), is more problematic. Since a yield curve is determined by all the U.S. STRIPS bonds, the movement of the yield curve should be represented by the movements of all the bond prices. But the movements of all the bond prices are not independent of each other. They have to be correlated. The following empirical evidence may suggest how the yield curve movements may be best specified.

25.2.1 Lognormal Versus Normal Movements

The movements (often referred to as the *dynamics*) of each interest rate of the spot yield curve can be specified as we have done for a stock. We can rewrite Eq. (1), replacing the stock price with a rate that is the yield to maturity of a zero coupon bond of a specific maturity “ t .” Thus we have:

$$dr = \mu(r, t)rdt + r\sigma dZ \quad (2)$$

When a t year rate is assumed to follow the process specified by Eq. (2), we say that the interest rate follows a *lognormal process* and Eq. (2) is called a *lognormal model*. In comparing Eq. (2) with Eq. (1), note that the drift term of the interest rate model is any function of the short-term interest rate r and time, while the lognormal model for stock tends to assume that the instantaneous stock return is a constant number. Therefore, the research literature of interest rate models has somewhat abused the language in calling Eq. (2) a lognormal model. The important point is that, in a lognormal process, the volatility term is proportional to the interest rate level $r(t)$. When the interest rate level is high, we experience high interest rate volatility. When the interest rate level is low, we experience low interest rate volatility.

There is an alternative specification of the interest rate process, which research literature calls the *normal process*. In the normal process, the volatility is independent of the interest rate level, and it is given below:

$$dr = \mu(r, t)dt + \sigma dz \quad (3)$$

Equation (3) is called the *normal model*. Note that the distinction made between the lognormal model and the normal model depends only on the volatility term and not on the drift term. For a normal model, the interest rate fluctuates with a volatility independent of the interest rate level over a short time interval. For a lognormal model, the interest rate has a volatility related to the interest rate level, in particular, when the volatility becomes arbitrarily small as interest rate level approaches zero. This way, the interest rates can never become negative. And a lognormal process is written as:

$$\frac{dr}{r} = \mu(r, t)dt + \sigma dZ \quad (3a)$$

Based on historical observations, the yield curve movements have been shown to be both normal and lognormal depending on the interest rate levels. Which model is more appropriate to describe interest rate movements, the normal or lognormal model? We need to evaluate the model from an empirical perspective. Using U.S. historical interest rates, the squared change of the interest rate over a 1 month period could be plotted against the interest rate level. Then we can see that the interest rate volatility has no relationship between the interest rate levels. If there were a positive relationship, we would see the higher volatility values related to higher interest rates. This result is consistent with Cheyette (1977), where he shows that the positive correlation between the interest rate volatility and the interest rate level is weak when the interest rate level is below 10%. However, when interest rate level was high in the late 1970s and early 1980s, the interest rate volatility was also high then, showing positive correlations only during that period.

Table 1 Correlation matrix of the interest rates

	0.25	0.5	1	2	3	5	7	10	20	30
0.25	1.000	0.936	0.837	0.701	0.630	0.533	0.443	0.377	0.087	0.083
0.5	0.936	1.000	0.938	0.832	0.770	0.675	0.587	0.509	0.224	0.154
1	0.837	0.938	1.000	0.940	0.895	0.816	0.731	0.654	0.379	0.291
2	0.701	0.832	0.940	1.000	0.989	0.950	0.898	0.832	0.573	0.426
3	0.630	0.770	0.895	0.989	1.000	0.980	0.945	0.887	0.649	0.493
5	0.533	0.675	0.816	0.950	0.980	1.000	0.982	0.946	0.736	0.595
7	0.443	0.587	0.731	0.898	0.945	0.982	1.000	0.976	0.821	0.670
10	0.377	0.509	0.654	0.832	0.887	0.946	0.976	1.000	0.863	0.750
20	0.087	0.224	0.379	0.573	0.649	0.736	0.821	0.863	1.000	0.867
30	0.083	0.154	0.291	0.426	0.493	0.595	0.670	0.750	0.867	1.000

25.2.2 Interest Rate Correlations

We have discussed the dynamics of interest rates. Now, let us consider the co-movements of interest rates. Do interest rates move together in steps, such that they all rise or fall together?

While the yield curve in principle can take many shapes historically, all the interest rates along the yield curve are positively correlated.

But the interest rates do not shift by the same amount. The co-movements of the interest rates can be investigated by evaluating the correlations of the interest rates, as presented in Table 1.

The results show that all the correlations are positive, which suggests that all the interest rates tend to move in the same direction. The long rates, which are the interest rates with terms over 10 years, are highly correlated, meaning that the segment of the yield curve from a 10- to 30-year range tends to move up and down together. The interest rates that are closer together along the yield curve have higher correlations.

25.2.3 Term Structure of Volatilities

Interest rate volatility is not the same for all interest rates along the yield curve. By convention, based on the lognormal model, the uncertainty of an interest rate is measured by the annualized standard deviation of the proportional change in a bond yield over a time interval (dt). For example, if the time interval is a 1 month period, then dt equals 1/12 year. This measure is called the interest rate volatility and it is denoted by $\sigma(t, T)$, the volatility of the T -th year rate at time t . More precisely, the volatility is the standard deviation of the proportional change in rate over a short time interval, and it is given by:

Table 2 Historical term structure of volatilities; $\sigma(\Delta r(t)/r(t)) \cdot \sqrt{12}$

0.25	0.5	1	2	3	5	7	10	20	30
0.1906	0.1908	0.1872	0.1891	0.1794	0.1632	0.1487	0.1402	0.1076	0.1137

$$\sigma(t, T) = Std\left(\frac{\Delta r(t, T)}{r(t, T)}\right) / \sqrt{\Delta t} \quad (4)$$

where $r(t, T)$ is the yield-to-maturity of the zero-coupon bond with time-to-maturity T at time t and $Std(\cdot)$ is a standard deviation over dt . We can relate Eq. (4) to (3a) by the following algebraic manipulations. For a small time step, Eq. (3a) can be written as:

$$\frac{\Delta r(t, T)}{r(t, T)} \cong \mu \Delta t + \sigma(t, T) \Delta Z$$

For sufficiently small Δt , we have:

$$\sigma\left(\frac{\Delta r(t, T)}{r(t, T)}\right) \cong \sigma(t, T) \sqrt{\Delta t}$$

Rearranging the terms, we can express σ as Eq. (4) requires. Similarly, based on the normal model, the term structure of volatilities is given by

$$\sigma(t, T) = \sigma(\Delta r(t, T)) / \sqrt{\Delta t} \quad (5)$$

The relationship of the volatilities with respect to the maturity is called *the term structure of volatilities*. The interest rate volatilities can be estimated using historical monthly data ($\Delta t = 1/12$). Below is the standard deviation of the rates for 0.25, 0.5, 1, 2, 3, 5, 7, 10, 20, 30 years.

The historical term structure of volatilities shows that the short-term rates tend to have higher volatilities than the long-term rates, falling from 19.06% for the 0.25-year rate to 11.37% for the 30 year rate. The empirical results suggest that we cannot think of interest rate volatility as one number. The volatility has to depend on the term of the interest rate in question (Table 2).

25.2.4 Mean Reversion

Thus far the discussion focuses on the volatility term of the dynamics of the interest rates. Now we investigate the drift term of interest rate movements. Research tends to argue that the yield curve cannot follow a random walk like a stock, as in Eq. (1). The yields of the Treasury bonds cannot rise and fall with the expected drift, yet to be constant or at a certain fixed proportion to the interest rate level. Since the nominal interest rate, which is what we are concerned with here,

is decomposed into the real interest rate and the expected inflation rate as stated in the Fisher equation, the movements of the nominal rates can be analyzed by considering the movements of the real rates and the inflation rate. One may argue that the real rate cannot follow a random walk because the real rate is related to all the individuals' time value of money in real terms. We tend to think the real interest rate is quite stable and that the real rate does not follow a random walk like a stock. To the extent that we believe the government seeks to control the inflation rate of an economy, the inflation rate cannot follow a random walk either. Therefore, we cannot assume that the (nominal) interest rate follows a random walk.

One may conclude that the interest rates tend to fall when the interest rates are high. Conversely, the interest rates tend to rise when interest rates are low. This is a somewhat imprecise description of a yield curve behavior, but we will provide a more precise description of this behavior later in the chapter, where we will provide alternative interest rate models in specifying this behavior. Research literature calls the dynamics that describe this behavior of interest rates a *mean reversion process*.

25.3 Equilibrium Models

Interest rate models seek to specify the interest rate movements such that we can develop a pricing methodology for an interest rate option.

25.3.1 The Cox-Ingersoll-Ross Model

The Cox, Ingersoll, and Ross (CIR) (1985) interest rate model is based on the productive processes of an economy. According to the model, every individual has to make the decision of consuming and investing with their limited capital. Investing in the productive process may lead to higher consumption in the following period, but it would sacrifice consumption today. The individual must determine the optimal trade off.

Now assume that the individual can also borrow and lend capital to another individual. Each person has to make economic choices. The interest rates reach the market equilibrium rate when no one needs to borrow or lend. The model can explain the interest rate movements in terms of an individual's preferences for investment and consumption as well as the risks and returns of the productive processes of the economy.

As a result of the analysis, the model can show how the short-term interest rate is related to the risks of the productive processes of the economy. Assuming that an individual requires a premium on the long-term rate (called term premium), the model continues to show how the short-term rate can determine the entire term structure of interest rates and the valuation of interest rate contingent claims.

The CIR model

$$dr = a(b - r)dt + \sigma\sqrt{r} dZ \quad (6)$$

Cox et al. (1985) offer one of the earlier attempts at modeling interest rate movements. The proposed equilibrium model extends from economic principles of interest rates. It assumes mean reversion of interest rates. As we have discussed in the previous section, mean reversion of interest rates means that when the short-term interest rate (r) is higher than the long-run interest rates (b), the short-term rate would fall adjusting gradually to the long-run interest rate. Conversely, when the short-term interest rate is lower than the long-run interest rate, the short-term rate would rise gradually to the long-run interest rate. Note that the long-run interest rate is not the long-term interest rate. Long-term interest rates continuously make stochastic movements, while the long-run interest rate is a theoretical construct, hypothesizing that the economy has a constant long-run interest rate that interest rates converge to over time. The constant (a) determines the speed of this adjustment. If the constant (a) is high/low, the adjustment rate to the long-term rate would be high/low. The CIR model is a lognormal model since the interest rate volatility is positively related to the interest rate level. The classification of lognormal and normal is based on the uncertain movement of the interest rate over a short period of time as described above.

25.3.2 The Vasicek Model

The second model is called the Vasicek model (1977). This model is similar to the CIR model such that the model assumes that all interest rate contingent claims are based on short-term interest rates. The only difference is that the volatility is not assumed to be dependent on the interest rate level, and therefore it is a normal model.

The Vasicek model

$$dr = a(b - r)dt + \sigma dZ, (a > 0) \quad (7)$$

These models assume that there is only one source of risk and the models are referred to as *one-factor models*. This assumption implies that all bond prices depend on the movements of the rate (r), and that all bond prices move in tandem because of their dependence on one factor. At first, this assumption seems to be unrealistic because, as we have discussed, the yield curve seems to have many degrees of freedom in its movements, and therefore, how can we confine our yield curve to exhibit a one-factor movement?

25.3.3 The Brennan and Schwartz Two-Factor Model

For many purposes the one-factor model may not be appropriate to use as valuation models. An interest rate spread option is one example that a one-factor model may not be adequate to value. The values of some securities depend on the changing

interest rate spreads between the 2 year rate and the 10 year rate. The one-factor model assumes that all the interest rates that move in tandem would eliminate the risk of the spread between the 2 year and the 10 year rates.

One extension asserts that all the bond prices of all maturities are generated by the short-term interest rate and a long-term rate – the long-term rate being the consol bond, which has no maturity and whose rate represents the long-term rate. Different versions of the two-factor models have been proposed in the following papers: Brennan and Schwartz (1979, 1982), Richard (1978), and Longstaff and Schwartz (1992). The Brennan and Schwartz model is given below:

$$\begin{aligned}dr &= a_1 + b_1(l - r)dt + r\sigma_1 dZ \\dl &= l(a_2 + b_2r + c_2l)dt + l\sigma_2 dW\end{aligned}\tag{8}$$

where r is the short-term rate and l is the consol rate, and where a consol bond is a bond that pays a fixed coupon periodically into the future on a notional amount with no maturity. σ_1 and σ_2 are the standard deviations of the short-term and consol rate, respectively. dZ and dW represent the risks which may be correlated. All the parameters a_1 , b_1 and a_2 , b_2 , c_2 are estimated from the historical data.

25.4 Arbitrage-Free Models

From the standard economic theory perspective, arbitrage-free modeling takes a departure from the CIR approach. The main point of the departure is sacrificing the economic theory in providing a model of the term structure of interest rates for a more accurate tool for valuing securities. Since the yield curve measures the agents' time value of money, the standard economic theory relates the interest rate movements to the dynamics of the economy. By way of contrast, arbitrage-free modeling assumes the yield curve follows a random movement much like the model used to describe a stock price movement. We can show that stock prices are assumed to be random and such an assumption does not incorporate the modeling of the agent's behavior and the economy.

25.4.1 The Ho–Lee Model

Ho and Lee (1986, 1990, 2004) takes a different approach in modeling yield curve movements as compared to CIR and Vasicek. The arbitrage-free interest rate model uses the relative valuation concepts of the Black–Scholes model Black (1995). This concept of relative valuation becomes a more complex concept to accept in the interest rate theory. Arbitrage-free modeling, like the Black–Scholes model, argues that the valuation of interest rate contingent claims is based solely on the yield curve. Economic research focuses on understanding the inferences made from the yield curve shape and its movements. The arbitrage-free model omits all these fundamental issues, apparently ignoring part of the economic theory behind interest rate

research. The model assumes that the yield curve moves in a way that is consistent with the arbitrage-free condition.

Let us assume that there is a perfect capital market in a discrete time world. But this time, the binomial model is applied to the yield curve movements. We assume:

- (1) Given the initial spot yield curve, the binomial lattice model requires that the yield curve can move only up and down.
- (2) The one period interest rate volatility (the instantaneous volatility) is the same in all states of the world.
- (3) There is no arbitrage opportunity in any state of the world (at any node point on the binomial lattice).

Assumption (1) is a technical construct of the risk model. Assumption (2) is made simply for this example. This assumption can be altered. Assumption (3) is the most interesting and important, called the “arbitrage-free condition.” This arbitrage-free condition imposes constraints on the yield curve movements.

Thus far it seems that the extension is directly from the Black–Scholes model. But there is one problem: interest rate is not a security. We cannot buy and sell the one-period rate, though we can invest in the rate as the risk-free rate. Moreover, we cannot use the one-period rate to form an arbitrage argument as the Black–Scholes model does with stock, since the one-period rate is the risk-free rate, which obviously cannot be the “underlying asset” as well. In equity option, the stock is both the underlying instrument as well as the risk source or the risk driver.

- A. **Arbitrage-free hedging:** The conceptual extension of the interest rate arbitrage-free model from the Black–Scholes model is to introduce the short-term interest rate as the risk source (or risk drive or state of the world). The Black–Scholes model’s risk neutral argument requires an underlying security and the risk-free rate. However, in the interest rate model, the risk-free rate is the risk source. One condition we want to impose on the interest rate movement is arbitrage-free, that is, the interest rate movements do not allow any possible arbitrage opportunity in holding a portfolio of bonds at any time. Research shows that the interest rate movements are arbitrage-free if the following two conditions hold (Harrison and Kreps 1979): (1) all the bonds at any time and state of the world have a risk-neutral expected return of the prevailing one period rate and (2) any bond on the initial yield curve has the risk-neutral expected return of the one-period interest rate of the initial yield curve. That is, for an interest rate movement to be arbitrage-free, there must be a probability assigned to each node of a tree such that all interest rate contingent claims have an expected “risk-free return,” which is the one-period rate. Note that this probability is the “risk neutral,” where the market probability can be quite different.
- B. **Recombining condition:** For tractability of the model, we require the discount function to recombine in a binomial lattice. This requirement is similar to the Black–Scholes model. Namely, the yield curve making an up movement and then a down movement must have the same value as the yield curve that makes a

down movement and then an up movement. The difference between the yield curve movement and the stock movement is that we need the entire discount function (or the yield curve), and not just one bond price, to be identical when they recombine.

Under these restrictions, we can derive all the possible solutions. Let us consider the simplest solution for us to gain insight into these arbitrage-free models. Suppose the spot yield curve is flat. The spot curve can shift in a parallel fashion up and down. The binomial lattice represented is called “normal” (or arithmetic) because the parallel shift of the curve is a fixed amount and not a proportion of the value at the node. The movements of the discount function can be represented by the binomial movements.

The purpose of the arbitrage-free model is not to determine the yield curve from any economic theory or to hypothesize that the yield curve should take on particular shapes. The arbitrage-free model takes the yield curve (or the discount function) as given, and then hypothesizes the yield curve (or the discount function) movements in order to relatively value other interest rate derivatives. Using a dynamic hedging argument similar to the Black–Scholes model, the argument shows that we can assume the local expectation hypothesis to hold: the expected return of all the bonds over each time step is the risk-free rate, the one-period interest rate.

The Ho–Lee model is similar to the Vasicek model in that they are both normal models. The main difference of course is that the Ho–Lee model is specified to fit the yield curve, whereas the Vasicek model is developed to model the term structure of interest rates. For this reason, the Vasicek model has the unobservable parameter called term premium, and the yield curve derived from the Vasicek model is not the same as the observed yield curve in general. Unlike the Vasicek model, the arbitrage-free interest rate model does not require the term premium, which cannot be directly observed. Instead, the arbitrage-free interest model only requires the given observed yield curve to value bonds. Hence, the theoretical bond prices would be the same as those observed.

Specifically, let the initial discount function, prices of zero-coupon bonds with a face value of \$1 and with maturity T , be denoted by $P(T)$. The discount function $P(T)$, for example, may be observed from the STRIPS market. The yield of the bond $P(T)$ is denoted by $r(T)$. Let σ be the volatility of the interest rate. Interest rate volatility may be estimated from historical data. Then the price of a one-period bond $P_i^n(1)$ in time n and state i on the binomial lattice is given by:

$$P_i^n(1) = 2 \left[\frac{P(n+1)}{P(n)} \right] \cdot \frac{\delta^i}{(1 + \delta^n)} \quad (9)$$

where

$P_i^n(1)$ = a one - period bond price at time period
 n and state i ,

$$\delta = e^{-2r(1)\sigma},$$

$$\sigma = \text{Std.} \left(\frac{\Delta r(1)}{r(1)} \right).$$

$-0.5 \ln \delta$ is the standard deviation of the change of the interest rate over each step size, while σ is the standard deviation of the proportional change of the interest rate.

While Eq. (9) provides the bond price for one period at any state i and time n , the model also has closed form solutions for bonds with any maturity at any node point on the lattice.

The basic idea of the derivation is quite simple, though the manipulation of the algebra is somewhat laborious. To derive the model, we need to determine the close form solution for $P_i^n(T)$, the price of a T year zero-coupon bond, at time n and state i , such that, under the risk-neutral probability 0.5, the expected return of a zero-coupon bond with any maturity, at any node point, equals the one-period risk-free rate. That is:

$$P_i^n(T) = 0.5 P_i^n(1) \{P_i^{n+1}(T-1) + P_{i+1}^{n+1}(T-1)\} \quad (10)$$

and we need to satisfy the initial observed yield curve condition:

$$P(T) = 0.5 P(1) \{P_0^1(T-1) + P_1^1(T-1)\} \quad (11)$$

The above equations hold for any i , n , and T . Then the model is assumed to be arbitrage-free in that all bonds have the expected returns and the bond pricing consistent with the initial spot yield curve (or the discount function $P(T)$).

Equation (9) specifies the one-period bond price (and hence the one-period interest rate) on each node of the binomial lattice. For this reason, we say that the model is an interest rate model, as the model specifies how the short-term interest rate movements are projected into the future.

We can show that once we can specify the one-period rate on a lattice, we can determine all the bond prices at each node point on the lattice by a backward substitution procedure similar to that used by the Black-Scholes model.

We can define the one period rate to be

$$r_i^n(1) = -\ln P_i^n(1) \quad (12)$$

Using Eq. (12), we see that the $r_i^n(1)$ can be expressed in three terms:

$$r_i^n(1) = \ln \frac{P(n)}{P(n+1)} + \ln \left(0.5 \left(\delta^{-(n/2)} + \delta^{n/2} \right) \right) + \left(\frac{n}{2} - i \right) \ln \delta \quad (13)$$

The first term is the one-period forward rate. That means that under the arbitrage-free interest rate movement model, we can think of the movement of the short-term rate as based on the forward rates. When there is no interest rate uncertainty, ($\delta = 1$), both the second and third terms are equal to zero, and therefore, the one-period forward rates define the future spot rate arbitrage-free movements.

The last term specifies the cumulative upward and downward shifts of the rates after n periods. It is important to note that the sizes of all the shifts are the same, $\ln \delta$.

That means the interest rate risk is independent of the level of interest rate, and the interest rate follows a normal distribution.

The second term is more difficult to explain as well as important. Let us consider a 2 year bond. Assume that the yield curve is flat at 10%. The bond price is therefore 0.826446. After 1 year, the interest rate shifts to 20% or 0% with equal probability, just to exaggerate the problem a little bit. The expected price of the bond is now

$$0.91667 \left(\frac{1}{2} \times \left(\frac{1}{1.2} \right) + \frac{1}{2} \times \left(\frac{1}{1.0} \right) \right).$$

The expected return of the bond over the first year is $10.9095 (= (0.91667 / 0.826446) - 1)$. Therefore, even with a yield curve that is flat at 10%, the yield curve makes the shifts of up or down with the same probability, and the expected return of the bond exceeds 10%. The reason is straightforward: When the interest rate moves, the bond price does not move in step with the interest rates. This is simply a matter of bond arithmetic of the yield calculation, where the yield is in the denominator. We can show that bonds have positive convexity. When the yield curve makes a parallel shift up or down with equal probability, the expected bond price is higher than the prevailing bond price. After all, it is the positive convexity of a bond that motivates the barbell trades.

Since bonds have positive convexity, if the interest rate shifts up or down by the same amount (with equal probability) relative to the forward rate, the expected returns of the bonds would exceed the one-period interest rate. To maintain the arbitrage-free condition, such that the local expectation hypothesis holds, we require the interest rate to shift higher in both up and down movements, so that the expected bonds' returns are equal to the one-period interest rate. That is, the interest rate movements must be adjusted upwards to correct for this convexity effect. This correction is the second term. Note that the second term, the convexity adjustment term, increases with the volatility as one may expect.

Thus far, we have discussed a set of interest rate models that exhibit normal distributions, which does not reflect the relationship between the interest rate uncertain movements and the interest rate levels. The lognormal model ensures that the interest rate uncertain movement increases or decreases with interest rate level. In particular, when interest rates continue to fall, the interest rate movement will continue to become smaller. In this case, the interest rates cannot become negative, while the normal model often has scenarios where the interest rates can become negative. An example of a lognormal model is the Black–Derman–Toy model.

25.4.2 The Black–Derman–Toy Model

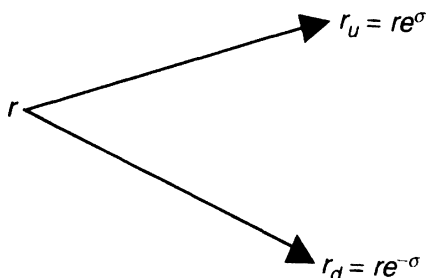
The Black–Derman–Toy (BDT) (1990) model is a binomial lattice model. This model assumes that the short-term interest rate follows a lognormal process. The model does not have a closed form solution and can best be explained by describing the procedure to construct the short-term interest rate movements.

The Black–Derman–Toy model uses a recombining lattice to determine a log-normal interest rate model. Further, the model can take the initial term structure of interest rate as input, as well as the term structure of volatilities, as in the extended Ho–Lee model. The model is specified by an iterative construction that can be best illustrated with an example:

As inputs to the model, we begin with the given term structure of interest rates and term structure of forward volatilities:

Maturity (years)	1	2	3	4	5
Yield (%)	6.0	7.0	8.0	9.5	10.0
Forward volatility (%)	15.0	14.0	13.0	11.0	

On the lattice, initially we have a one-period rate, say, 6%. The lognormal model is determined by the following random walk at a node:



Note that, using the definition of r_u and r_d , we know

$$r_u = r_d e^{2\sigma}. \quad (14)$$

Step 1. Construct the lowest short-term rate for each period in the lattice.

These rates are r , $r \cdot \exp[-\sigma(1)]\mu(1)$, $r \cdot \exp[-\sigma(2)]\mu(2)$. Note that we do not know μ , the only parameter unknown at this point.

Step 2. Specify the short-term rates at all the nodes using Eq. (14).

We need to iteratively calculate the rate r_u , applying Eq. (14) repeatedly.

Step 3. Determine μ by a “bootstrap” approach.

Search for the value $\mu(1)$ such that a 2 year bond, given by the discount function $P(T)$, can be priced according to the market. Then, we determine $\mu(2)$ such that $\mu(2)$ can price the 3 year bond exactly according to the observed (or given) 3 year bond price. This iterative procedure, called the bootstrap approach, can determine the lattice as desired.

We calculate the short rates by following the BDT procedure given the yields and the instantaneous forward volatilities in the table above.

25.4.3 The Hull–White Model

The Hull–White model (1990, 1993) is a normal model that has an explicit term to capture the mean reversion of interest rates. It is similar to the Vasicek model with the difference of being arbitrage-free. This approach enables the model to capture the

term structure of volatilities by adjusting the adjustment rate of the short-term rate to the long-term equilibrium rate. The lattice model they propose is not a binomial model but a trinomial model. The trinomial model enables the model to adjust for the speed of adjustment and it can be constructed such that the model has no negative interest rates in all scenarios.

The Hull–White model can also be extended to a two-factor model (1994) that is arbitrage-free in a form similar to the Brennan and Schwartz model. Specifically, the model is specified by two simultaneous equations:

$$dr = [\theta(t) + u - ar]dt + \sigma_1 dW \quad (15)$$

$$du = -budt + \sigma_2 dZ \quad (16)$$

In this case, the short-term rate makes partial adjustments to the long-term rate, while the long-term rate follows a random movement. Using normal model properties, these models can derive closed form solutions for many derivatives in the continuous time formulation.

References

- Black, F. 1995. Interest rates as options. *Journal of Finance* 50 (5): 1371–1376.
- Black, F., E. Derman, and W. Toy. 1990. A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal* 46: 33–39.
- Brennan, M.J., and E.S. Schwartz. 1979. A continuous time approach to the pricing of bonds. *Journal of Banking and Finance* 3: 135–155.
- . 1982. An equilibrium model of bond pricing and a test of market efficiency. *Journal of Financial and Quantitative Analysis* 17: 301–329.
- Cheyette, O. 1977. Interest rate models. In *Advances in fixed income valuation, modeling and risk management*, ed. Frank J. Fabozzi. New Hope: Frank J. Fabozzi Associates.
- Cox, J.C., J.E. Ingersoll Jr., and S.A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53: 385–407.
- Harrison, M.J., and D.M. Kreps. 1979. Martingales and arbitrage in multi-period securities markets. *Journal of Economic Theory* 20: 381–408.
- Ho, T.S.Y., and S. Lee. 1986. Term structure movements and pricing of interest rate contingent claims. *Journal of Finance* 41: 1011–1029.
- . 1990. Interest rate futures options and interest rate options. *Financial Review* 25 (3): 345–370.
- Ho, T.S.Y., and S.B. Lee. 2004. *The Oxford guide to financial modeling*. New York: Oxford University Press.
- Hull, J., and A. White. 1990. Pricing interest-rate-derivative securities. *Review of Financial Studies* 3 (4): 573–592.
- . 1993. One-factor interest-rate models and the valuation of interest-rate derivative securities. *Journal of Financial and Quantitative Analysis* 28 (2): 235–254.
- . 1994. Numerical procedures for implementing term structure model II: Two-factor models. *Journal of Derivatives* 2: 37–48.
- Longstaff, F.A., and E.S. Schwartz. 1992. Interest rate volatility and the term structure: A two factor general equilibrium model. *Journal of Finance* 47 (4): 1259–1282.
- Richard, S.F. 1978. An arbitrage model of the term structure of interest rates. *Journal of Financial Economics* 6: 33–57.
- Vasicek, O. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5: 177–188.

Chiuling Lu and Chou-Yen Wu

Abstract

In this chapter the history and the success of Real Estate Investment Trusts (REITs) and Mortgage-Backed Securities (MBS) in the US financial market are discussed. REITs had both great losses during financial crisis of 2008–2009 and the COVID-19 pandemic of 2020. They enhanced their financial positions after the financial crisis and their leverage ratios at the lowest records before near the pandemic. REITs have had a partial recovery at the end of 2020 and are more likely to recover in 2022. Both REITs and MBS are derived from real estate-related assets and are able to increase the liquidity on real estate investment. They also provide investors with the opportunity to diversify portfolios because real estate assets are relatively less volatile and less correlated to existing investment instruments. Therefore, REITs and MBS enhance the width and the depth of the financial market.

Keyword

REIT · MBS · Real Estate · Mortgage · FHA · VA · Fannie Mae · Ginnie Mae · Freddie Mac · Prepayment · Public Securities Association

JEL Classification

G10 · L85

C. Lu (✉)

Department of International Business, National Taiwan University, Taipei, Taiwan
e-mail: chiulinglu@ntu.edu.tw

C.-Y. Wu

Department of Accounting, Feng Chia University, Taichung, Taiwan
e-mail: chouywu@fcu.edu.tw

Abbreviations

COVID-19 coronavirus disease 2019

26.1 Introduction

The revolution in the American real estate market was enhanced by securitization. For real properties, public listed Real Estate Investment Trusts (REITs) create tradable and standardized securities for individuals and institutional investors while providing alternative investments for diversification. Through the capital market, real estate practitioners have more reliable funds and are no longer limited to bank loans. For real estate-related loans, Mortgage-Backed Securities (MBS) establish the capital conduit linking borrowers and lenders directly and by-passing financial intermediaries. MBS also release the burden of the bank from holding long-term mortgage debt and bearing the credit risk. In addition, MBS create the secondary market of mortgage debt and provide an alternative investment.

Although REITs and MBS Mortgage-Backed Securities (MBS) are generated under different backgrounds and developed under different circumstances, they play a significant role in real estate financing. Nevertheless, challenges in real estate securitization have been important issues for the past six decades and will continue to do so in the future. The impact of the COVID-19 pandemic on many economic activities leads to big losses in operating income of REITs from rents and occupancies.

In this chapter, the development of REITs and MBS Mortgage Backed Securities (MBS) for the past 60 years in the USA is described. In addition, the empirical findings in the literature are also examined.

26.2 The REIT Background

A REIT is a creation of the federal tax code that permits an entity to own real properties and mortgage portfolios. REITs incur no corporate tax on transfers of profits to holders of beneficial interest given certain provisions within the Internal Revenue Code are met. To qualify as a REIT for tax purposes, the trust must satisfy many requirements including asset, income, distribution, and ownership restrictions. For example, a REIT must have a minimum of 100 shareholders; invest at least 75% of the total assets in real estate assets; derive at least 75% of gross income from rents, or interest on mortgages on real property; and pay at least 90% of the taxable income in the form of shareholder dividends. Basically, if any company fails to qualify as a REIT, the company cannot again elect to be taxed as a REIT until 5 years after the termination date. To know REITs better, this study begins with the origin of REITs, then discusses the related regulation changes and current development, and finally examines relevant accounting and financial issues.

REITs were created by the US Congress in 1960 to enable small investors to become involved in real estate development, which was previously limited to the affluent. However, for the first three decades, REITs were recognized as passively managed firms and were not competitive with real estate limited partnerships. Until the Tax Reform Act of 1986, REITs were empowered to not only own, but also operate and manage their own assets. In addition, the Act reduced tax shelter opportunities for real estate partnerships. Thereafter, it was possible for REITs to be self-managed rather than managed by external advisors, and became more attractive to investors. For the distribution rule, REITs were required to distribute at least 95% of taxable income as dividends. However, in 1999, the REIT Modernization Act changed the minimum requirement to 90% which is consistent with the rules from 1960 to 1980. Today, the US REIT approach has been the model for 39 countries.

According to the National Association of REIT (NAREIT), there are over 800 REITs and 219 of them were traded on the NYSE, AMEX, and NASDAQ during 2019 while only 34 REITs were traded during 1971. The market capitalization and the number of publicly traded REITs from 1971 to 2019 are illustrated in Fig. 1 Since its inception, REITs played a limited role in the capital market until the end of the 1980s. At the end of the 1980s, the combined effect of overbuilding, the savings and loan crisis, and the impact of the Tax Reform Act of 1986 led to the expansion of REITs. Investors realized that tradable and liquid real estate investment is crucial during recessions and REITs happened to fit these needs. However, the shock of Financial Crisis leads to a large decrease in market capitalization of REITs in 2008 and 2009. Figure 1 shows that the total market capitalization of REITs amounted to more than US\$1329 billion as of 31 December 2019. The mean of market capitalization also grows by years. For example, the mean of market capitalization of REITs

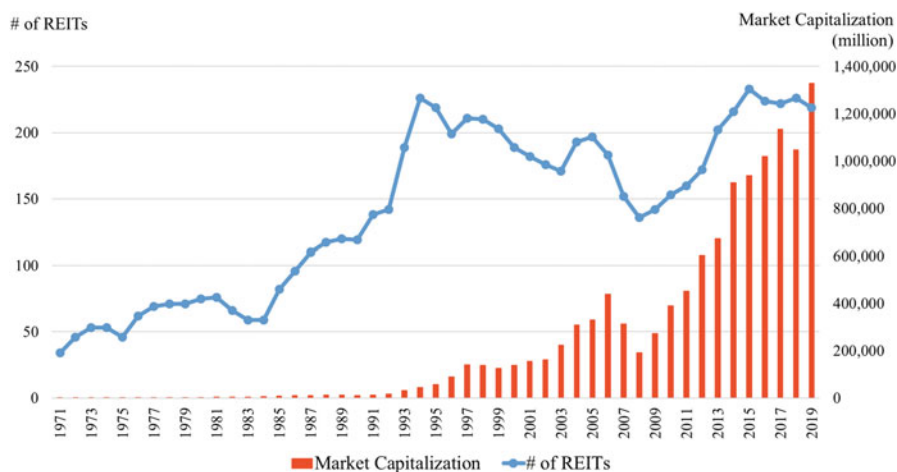


Fig. 1 REIT Equity Market Capitalization Outstanding Source: NAREIT

at the end of 2000 is US\$733.94 million and it increased to US\$6067.61 million at the end of 2019.

Because real estate maintains greater residual value than other assets such as computers or machinery, and real estate may appreciate at the same time, depreciation used in normal earnings measures resulted in underestimated cash flows for REITs. Bradley et al. (1998) observed that REITs' depreciation expenses are roughly equal to net income, and cash flow available for distribution is about twice the required payout. Consequently, agency problems caused by free cash flows arise (see Lu and Shen (2004)). In order to estimate cash flows and evaluate REITs' performance more accurately, Funds from Operations (FFO) is designed to be another supplemental measure relative to Earnings Per Share. The NAREIT defines FFO as net income excluding gains or losses from sales of property or debt restructuring, and adds back the depreciation of real estate.

A high dividend yield (7% on average in 2002) is one attraction for investors to invest in REITs. Figure 2 shows all Equity REITs' dividend payout ratios as a percentage of the FFO. At the end of 2000, only 78% of the FFO was distributed to shareholders. This ratio increased to 83% and 99% at the end of 2010 and at the second quarter of 2020, respectively. At the end of 2008, the equity REITs' dividend payout ratios were much high at 192% of FFO because FFO might be much low during the crisis. Basically, the dividend policy of REITs is quite different from that of nonfinancial firms (see Lee and Kau 1987) and is highly regulated by the IRS.

Compared to income-producing commercial real estate, REITs are financed on a more conservative basis. Generally, REITs finance their projects with about half debt and half equity. According to the NAREIT, the average debt ratio for equity REITs is 41.8% as of the fourth quarter of 2003 and about two-thirds of REITs with senior unsecured debt ratings are investment grade. Figure 3 shows that the average leverage ratio of REITs has decreased from the first quarter of 2009 to the second quarter of 2020, but no more than 40% since the end of 2010. Finally, the average debt ratio for equity REITs decreases to 33.8% in the third quarter of 2020. Figure 3 indicates that REITs strengthen their financial position after the financial crisis. The coverage ratios defined as dividing EBITDA by interest expenses has been less than 3 before the first quarter of 2009. It remained at 4–5 from the third quarter of 2014 to the second quarter of 2020. The valuation of REITs depends on several criteria including management quality, dividend coverage from FFO, anticipated growth in FFO, and economic outlook. Fortunately, because public REITs are traded every day, stock prices reflect real-time pricing. Therefore, a capital asset pricing model could be employed to calculate REITs' expected returns and systematic risk. According to previous research, REITs underperform the market on a nominal basis and earn fair returns on a risk-adjusted basis. Glascock and Hughes (1995) found that the REIT betas are consistently below the market (=1) and equal to 0.377 for the entire period (1972–1991). Figure 4 compares the dividend yield of REITs with a 10-year constant maturity treasury yield. On average, the REIT dividend yield and 10-year treasury yield were 6.52% and 6.19% from 1972 to January 2021, respectively. Figure 4 also indicates that the former was higher or slightly higher for most of time during the past 48 years, except 1980s. In addition, the difference

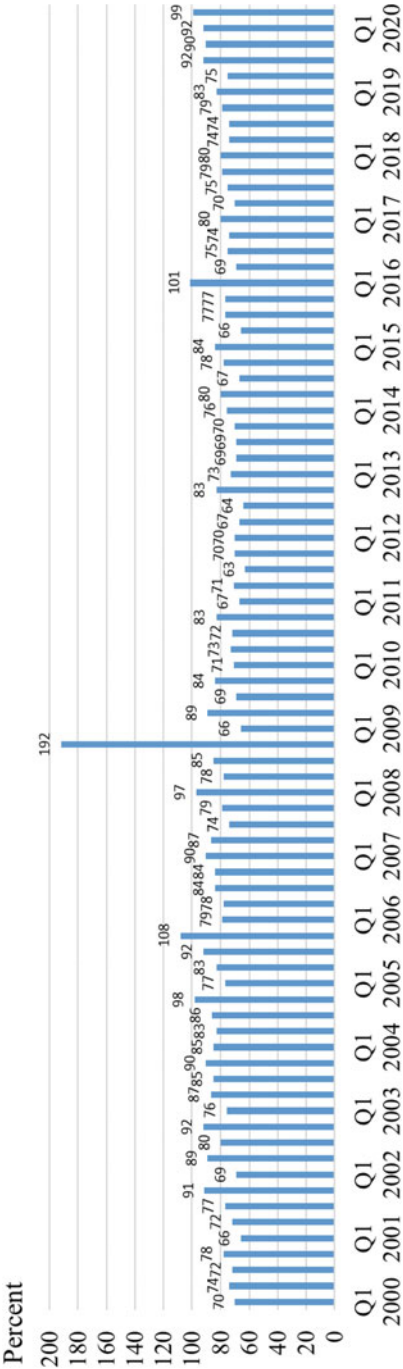


Fig. 2 FTSE Equity REITs – Payout Ratios: Dividends as a Percent of FFO (Adjusted Quarterly, Q1 2000–Q2 2020) Source: NAREIT

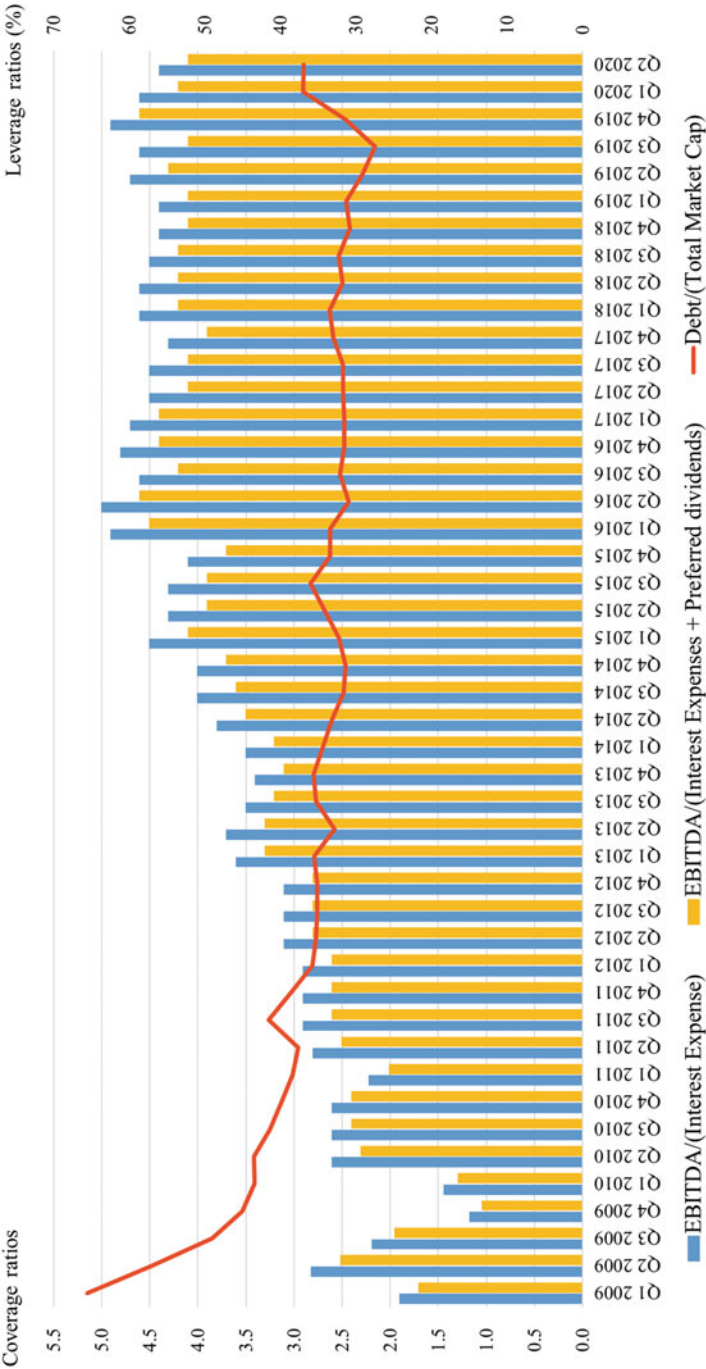


Fig. 3 Equity REIT Leverage and Coverage Ratios (Q1 2009–Q2 2020) Source: NAREIT

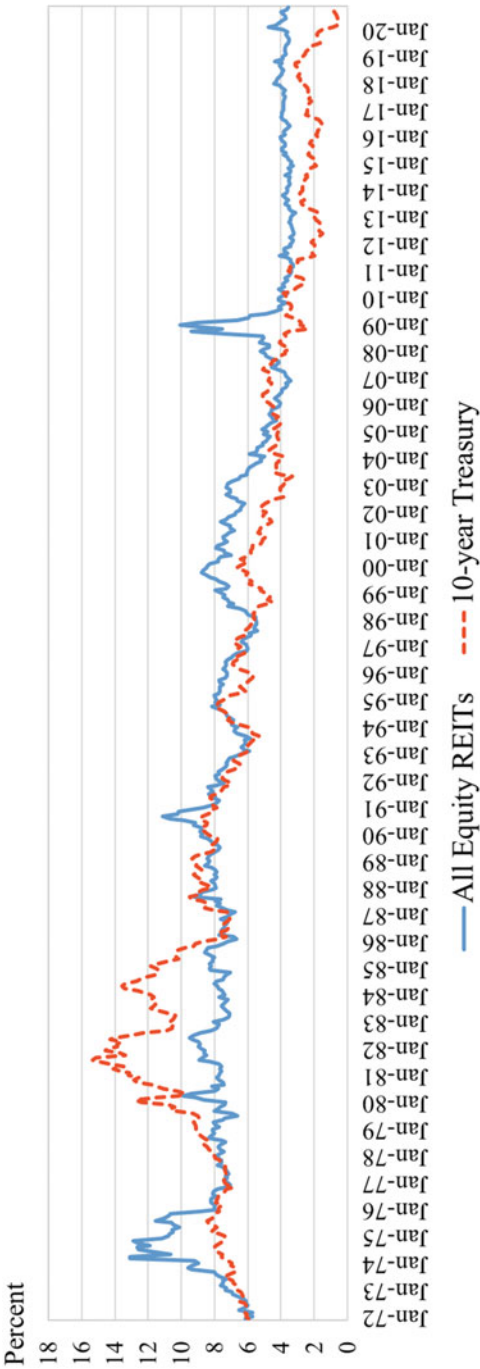


Fig. 4 NAREIT Equity REIT Dividend Yield vs. 10-year Constant Maturity Treasury Yield (January 1972–December 2020) Source: NAREIT, St. Louis Fed

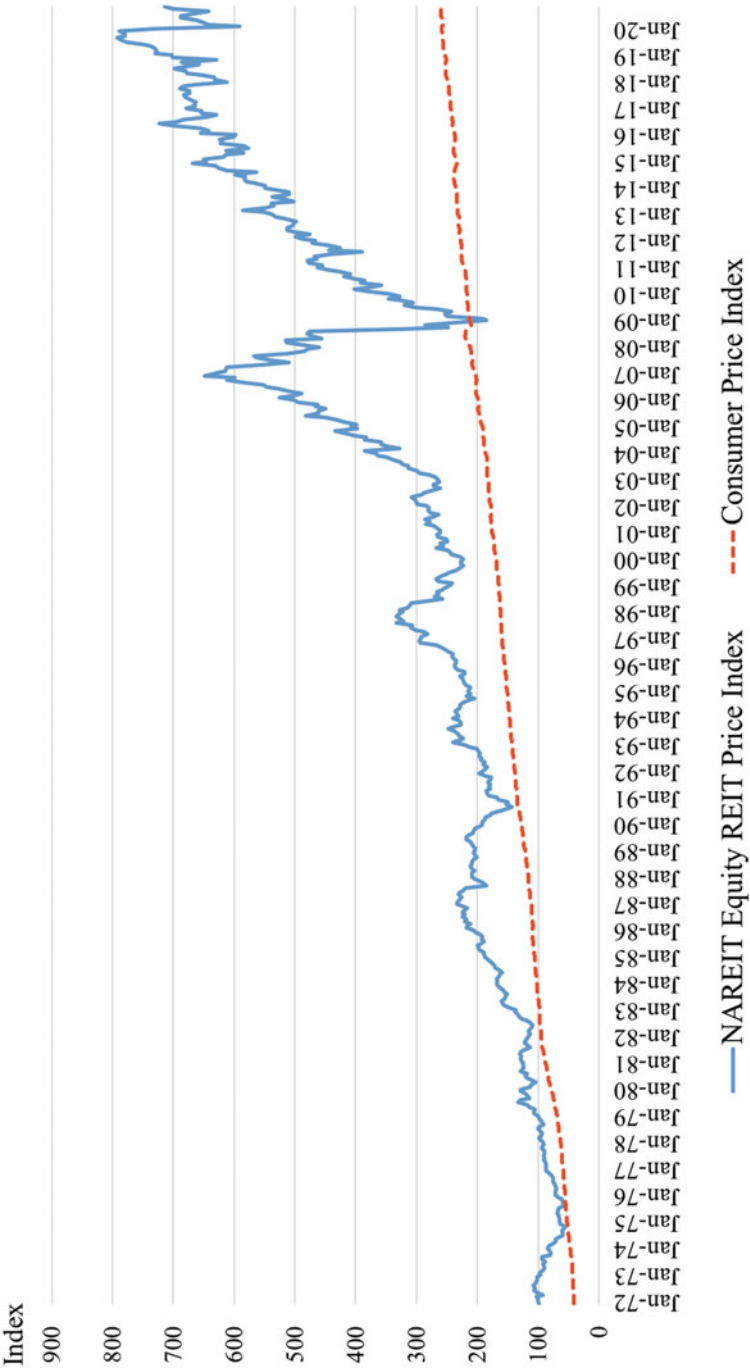


Fig. 5 NAREIT Equity REIT Price Index vs. Consumer Price Index (January 1972–December 2020) Note: NAREIT equity REIT price index is at December 1971 = 100; US consumer price index is not seasonally adjusted and at 1982–1984 = 100. Source: NAREIT, St. Louis Fed

between these two yields has increased in the 2000s. However, the dividend yield of REITs was much lower than a 10-year constant maturity treasury yield in 1980s. For example, REIT dividend yield was 7.96% on average during 1980s and the 10-year treasury yield was 10.60% on average in the same period. In general, real estate investment has been treated as a good way to hedge inflation. Fortunately, REITs investment still preserves this function. Figure 5 shows the trend of the NAREIT equity REIT price index versus the Consumer Price Index from 1972 to 2020. In the long run, the equity REIT price index was higher than the Consumer Price Index. Figure 5 also indicates that the difference between the NAREIT equity REIT price index and the Consumer Price Index were larger after 2000s. Furthermore, NAREIT Equity REIT Price Index was lower than Consumer Price Index during 2008 while NAREIT Equity REIT Price Index was much higher than Consumer Price Index during COVID-19 pandemic of 2020. Consequently, REITs are a viable inflation hedge instrument.

REITs have faced two tough periods in the past 15 years. With the severe COVID-19 pandemic in the USA, there were big losses in operating income of commercial REITs from rents, occupancies, and prices because of the implications of strict lockdown and social distancing. Akinsomi (2021) finds that the 3-month return performance of the FTSE EPRA NAREIT Americas index had the biggest losses as at 22 May 2020 related to other REIT markets. Compared to 2019, lodgings and retail have the largest losses in negative returns among the sectors of REITs. Diversified, office and residential REITs were the second losses in negative returns. Finally, data center REITs was the only one at a positive return during the pandemic. Schnure (2021) had a similar result that FTSE NAREIT All Equity REITs Index had a return performance at -37.6% year-to-date as of March 2020 and it stood at -7.4% year-to-date. REIT performance had a great difference in property sectors.

The outlook for 2021 remains uncertain, however, as office and residential REITs were prompted to have operational changes due to working-from-home. The data center and cell towers REIT, however, are expected to benefit during the pandemic. The reason for the recession is also important because it is associated with the recovery ahead (Schnure 2021). The pandemic recession of 2020 was caused by an external shock while the financial crisis of 2008–2009 was caused by an internal subprime mortgage crisis. REITs strengthen their financial position after the financial crisis. For example, the leverage ratio declined from 65.5% at the first quarter of 2009 to 31.4% at the end of 2019 (see Fig. 3). The leverage ratios of REITs before the COVID-19 pandemic were at the lowest on record. Schnure (2021) also indicates that REITs have held cash and securities and also committed credit lines to maintain high liquidity. They also have extended the maturity of debts to reduce the risks of having to refinance under adverse market conditions. Thus, Schnure (2021) thinks that REITs are more likely to recover in 2022. Moreover, an increase in demand for the digital economy leads to the closure of bricks and mortar retail locations; however, it probably promotes an operating change (i.e., online orders with curbside pickup) for the retails. Thus, it might promote an operating change for commercial REITs.

REITs have helped increase the liquidity of the real estate market and have become viable investments for diversification purposes by institutional and individual investors. It is anticipated that this industry will continue to expand and more countries will follow this track.

26.3 The MBS Story

The secondary mortgage market has evolved and grown in the USA for the last 3 decades. Mortgage backed securities (MBS) provide mortgage originators with liquidity and facilitate a geographic flow of funds from places with a surplus of savings to where home mortgages are needed.

The strong support of the federal government played the most important role in the development of the MBS market. Therefore, the government's sponsorship in the process, and then the market structure and participants are examined. The MBS pricing and related risk will also be illustrated.

26.3.1 The Special Contributions of the Government-Sponsored Enterprises

The three most important events in the evolution of the secondary mortgage market were the creation of Federal Housing Administration (FHA) in 1934, the chartering of the Federal National Mortgage Association (FNMA or Fannie Mae) in 1938, and the origination of Veterans Administration (VA) in 1944.

The FHA and the VA helped set up the mortgage underwriting standard and provided mortgage default insurance or guarantees. The FNMA was transformed into a privately owned and managed organization under the Housing and Urban Development Act of 1968. After that, government ownership was eliminated and the FNMA became solely owned by private investors. This Act also created the Government National Mortgage Association (GNMA, Ginnie Mae) to deal with subsidized mortgage purchases for special federal housing programs.

Though the secondary mortgage market based on pools of FHA or VA home mortgages was well established, a market for conventional loans did not exist. The Federal Home Loan Mortgage Corporation (FHLMC, Freddie Mac) chartered under Title III of the Emergency Home Finance Act of 1970 provided liquidity for conventional loans as well as for FHA-VA mortgages.

26.3.2 Market Participants

Basically, there are four entities involved in the operation of the secondary mortgage market. The first entity is the mortgage originator such as mortgage bankers, thrifts, and commercial banks. They perform loan underwriting and establish loan terms in the primary market, and then sell mortgages to replenish funds. The second entity

involved is the FHA and the VA, which perform credit enhancement functions by providing insurance or guarantees. The third entity in the process is the mortgage buyers. Prior to the mid-1950s, buyers were life insurance companies or thrifts. However, after the mid-1950s, both the FNMA and the FHLMC became the predominant purchasers. The FNMA and the FHLMC in turn created mortgage pools for securitization. The FNMA was divided into FNMA and GNMA in 1968. The fourth entity is the end investors such as REITs, pension funds, mutual funds, IRAs, life insurance companies, or even the mortgage originators themselves.

26.3.3 MBS Pricing

According to Bartlett (1989), MBS is a hybrid investment in which a portfolio holding of an MBS consists of one part a standard-coupon bond and one part a short call option. Since the homeowner has the right to prepay the mortgage at any time, the right of prepayment is regarded as a call option, which implies the MBS investor is in effect short the implied call. Therefore, Bartlett defines the MBS price as follows.

$$\text{MBS Value} = \text{Non-callable Bond Value} - \text{Call Option Value}$$

Unlike traditional debt securities, the cash flow of MBS is unpredictable due to unexpected delayed payment, prepayment, or default. Because borrowers hold options, not only is the expected maturity of MBS more difficult to estimate relative to the other straight bond investments, but also the exact timing and amount of the cash flow is unknown in advance. Therefore, the valuation of the MBS turns out to be more complicated. Actually, the valuation of the MBS is a concern because of the subprime crisis in 2007.

The factors related to the pricing of MBS include, but are not limited to, interest rate risk, default risk, risk of delayed payment, and prepayment risk. Prepayment ratios effect the predication of cash flow the most significantly. Therefore, several models have been developed to estimate this rate. Those models include the 12-year prepaid life (based on FHA data) assumption, constant prepayment rate (CPR) assumption, FHA prepayment experience, the Public Securities Association (PSA) model, and the econometric prepayment models (see Brueggeman and Fisher 2001). Kau et al. (1985, 1987, 1990a, b, 1992, 1993, 1995) have developed several models to analyze different mortgages and MBS. Research is continuing on prepayment estimation, but no conclusive model has been developed as of yet.

Figure 6 presents the outstanding volume of public and private bond market debt from 1980 to the third quarter of 2020. The outstanding level of mortgage-related debt was only US\$111.4 billion in 1980 and increased to US\$10.907 trillion in the third quarter of 2020. The outstanding level of corporate debt and US treasury bonds was US\$10.439 trillion and US\$20.368 in the third quarter of 2020, respectively. Obviously, MBS market-related securities were higher than the corporate bond and US Treasury bond markets. This trend indicates the success and need of the

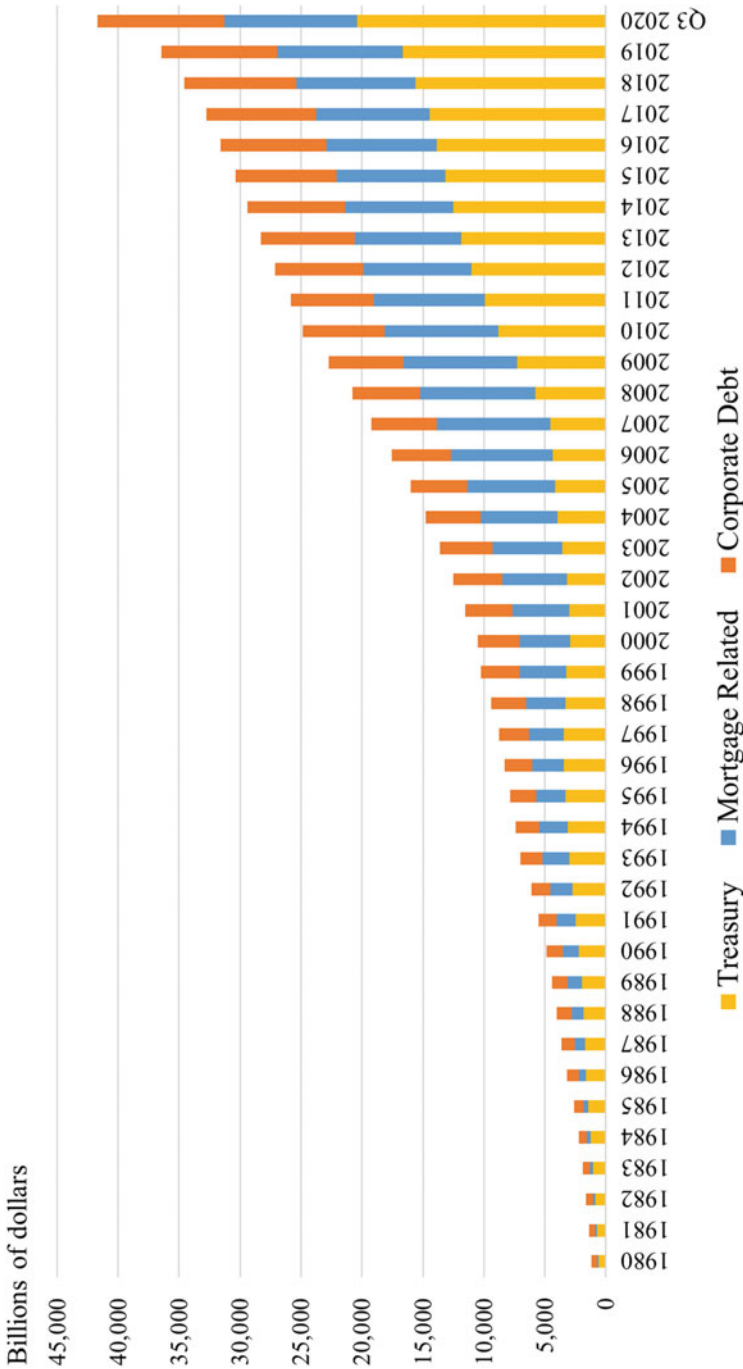


Fig. 6 Outstanding Level of Public and Private Bond Market Debt (Q1 1980–Q3 2020) Source: The Securities Industry and Financial Markets Association (SIFMA)

mortgage secondary market. The Commercial MBS yield spread is defined as the difference between AAA-rated 10-year CMBS and 10-year Treasuries. According to Morgan Stanley (2020a, b), AAA-rated CMBS yield spreads have tightened and retraced to pre-COVID-19 levels at around 80 basis points during the fourth quarter of 2020. During the quarter, AAA-rated CMBS spreads tightening 20–50 basis points because the 10-year Treasury rate rose by 23 basis points to 0.92%. Given similar risk levels, mortgage-related securities do provide investors with better alternative investments. The wider spreads remain for many CMBS thus many investors look to the CMBS market as a COVID-19 recovery opportunity. The Securities and Exchange Commission (SEC) and Financial Industry Regulatory Authority (FINRA) implicated new regulations (the Regulatory Notice 16–31) in December 2016 to mitigate the risks of CMBS through margin requirements for agency transactions. With the shock of COVID-19 pandemic, Schnure (2021) indicates that the CMBS delinquency rate peaked in June and had decreased to 8.2% in November 2020. The decline at the end of 2020 was due to the decreases in lodging and retail. There is expected to be a 40% chance that the CMBS delinquency rate will slightly decline due to the improvement of lodging and retail in 2021. But office delinquencies probably rose due to work from home. Finally, the CMBS market is expected to be a K-shaped recovery and a significant difference of winners and losers between property types in 2021 (Morgan Stanley 2020a).

26.3.4 The Impact of Securitization on Financial Institutions

The process of mortgage securitization helps financial institutions to manage their asset portfolios, interest rate exposure, capital requirement, and deposit insurance premiums. Saunders and Cornett (2003) state that asset securitization provides a mechanism for financial institutions to hedge the interest rate risk. They point out that the process of securitization not only make asset portfolio of financial institutions more liquid, but also provide an important source of fee income. Saunders and Cornett (2004) indicate that by increasingly relying on securitization, banks and thrifts have begun to move away from being asset transformers to become asset brokers. Therefore, the differences between commercial banking and investment banking become diminished as asset securitization expanded.

26.4 Conclusion

Equity securities, debt instruments, and derivatives have become popular investment or hedging vehicles during the past century. However, real estate, which is the most conventional investment asset, lost favor to liquid and tradable securities among investors. Not until the 1970s did institutional investors start to show an interest in real estate in the USA (see Bernstein 2003). The creation of REITs and MBS has changed the way of real estate financing in the USA and also has facilitated investment in real estate market. REITs and MBS were developed to complete the

market and indeed fulfill the objectives of an affordable housing policy for government and that of an asset allocation for the portfolio management purpose. Finally, the COVID-19 pandemic recession was caused by an external shock of economy. REITs have strengthened their financial positions after the financial crisis of 2008–2009; hence they are expected recovery in 2022.

References

- Akinsomi, O. 2021. How resilient are REITs to a pandemic? The COVID-19 effect. *Journal of Property Investment and Finance* 39: 19–24.
- Bartlett, W.W. 1989. *Mortgage-backed securities: Products, analysis, trading*. 1st ed. Englewood Cliffs: NJ, Prentice Hall.
- Bernstein, P. 2003. Real estate matters. *The Journal of Portfolio Management: Special Real Estate Issue*: 1. <https://jpm.pm-research.com/content/29/5>.
- Bradley, M., D. Capozza, and P. Seguin. 1998. Dividend policy and cash-flow uncertainty. *Real Estate Economics* 26: 555–580.
- Brueggeman, W., and J. Fisher. 2001. *Real estate finance and investments*. 11th ed. New York: McGraw-Hill/Irwin.
- Glascok, J., and W.T. Hughes. 1995. NAREIT Identified exchange listed REITs and their performance characteristics: 1972–1990. *Journal of Real Estate Literature* 3: 63–83.
- Kau, J.B., D.C. Keenan, W.J. Muller III, and J.F. Epperson. 1985. Rational pricing of adjustable rate mortgages. *AREUEA Journal* 13: 117–128.
- . 1987. The valuation and securitization of commercial and multifamily mortgages. *Journal of Banking and Finance* 11: 525–546.
- . 1990a. Pricing commercial mortgages and their mortgage-backed securities. *Journal of Real Estate Finance and Economics* 3: 333–356.
- . 1990b. The valuation and analysis of adjustable rate mortgages. *Management Science* 36: 1417–1431.
- . 1992. A generalized valuation model for fixed-rate residential mortgages. *Journal of Money, Credit, and Banking* 24: 279–299.
- . 1993. Option theory and floating rate securities with a comparison of adjustable- and fixed-rate mortgages. *Journal of Business* 66: 595–617.
- . 1995. The valuation at origination of fixed-rate mortgages with default and prepayment. *Journal of Real Estate Finance and Economics* 11: 5–36.
- Lee, C.F., and J.B. Kau. 1987. Dividend payment behavior and dividend policy of REITs. *Quarterly Review of Economics and Business* 27: 6–21.
- Lu, C., and Y. Shen. 2004. Do REITs pay enough dividends? Unpublished working paper, Department of Finance, Yuan Ze University.
- Morgan Stanley. 2020a. *2021 Global strategy outlook: Keep faith in the recovery*.
- . 2020b. *Morgan Stanley investment funds – Global asset backed securities fund*.
- Saunders, A., and M.M. Cornett. 2003. *Financial institutions management: A risk management approach*. 4th ed, 733–768. New York: McGraw-Hill/Irwin.
- . 2004. *Financial markets and institutions: A modern perspective*, 645–667. New York: McGraw-Hill/Irwin.
- Schnure, C. 2021. 2021 Outlook for REITs and commercial real estate: Risk and resilience top 10 issues to watch. *NAREIT Research*.



Experimental Economics and the Theory of Finance

27

Haim Levy

Abstract

Experimental findings and in particular Prospect Theory and Cumulative Prospect Theory contradict Expected Utility Theory, which in turn may have a direct implication to theoretical models in finance and economics. We show growing evidence against Cumulative Prospect Theory. Moreover, even if one accepts the experimental results of Cumulative Prospect Theory, we show that most theoretical models in finance are robust. In particular, the CAPM is intact even if investors make decisions based on change of wealth, employ decision weights, and are risk-seeking in the negative domain.

Keywords

Certainty effect · Configural weights · Cumulative prospect theory · Decision weights · Expected utility · Markowitz stochastic dominance · Prospect stochastic dominance · Prospect theory · Stochastic dominance · Value function

27.1 Introduction

Theoretical models in finance are based on certain assumptions regarding the investors' characteristics and their investment behavior. In particular, most of these models assume rational investors who always prefer more than less consumption (money), and who maximize von Neumann–Morgenstern (1944) expected utility.

H. Levy (✉)
Hebrew University, Jerusalem, Israel
e-mail: mshlevy@mscc.huji.ac.il

The main models in finance that we relate to in this chapter are:

- (1) The Modigliani–Miller (1958) relationship between the value of the firm and its capital structure.
- (2) Black–Scholes (1973) option pricing.
- (3) Ross’s (1976) Arbitrage Pricing Model (APT).
- (4) The Sharpe–Lintner (1964 and 1965, respectively) Capital Asset Pricing Model (CAPM).
- (5) Stochastic Dominance – the various investment decision rules (for a review, see Levy 1992, 1998).
- (6) Market Efficiency – though recently some empirical studies reveal (short term) autocorrelations, most academic research still assumes that the market is at least “weakly efficient,” namely one cannot employ *ex-post* rates of return to establish investment rules that provide abnormal returns. Of course, if this is the case, there is no room for “technicians” and charterists who try to predict the market based on past rates of return. (For the market efficiency hypotheses see Fama 1965, 1991).

In this chapter, we analyze the impact of recent experimental finding, and particularly the implication of Prospect Theory (PT) (see Kahneman and Tversky 1979) (K&T), Cumulative Prospect Theory (CPT) (see Tversky and Kahneman 1992 (T&K)), and Rank-Dependent Expected utility RDEU) (RDEU) (see Quiggin 1982, 1993) on each of these subjects that are cornerstones in finance and in decision making under uncertainty.

The structure of this chapter is as follows: In Sect. 27.2, we deal with the main findings of PT and their implication regarding the above mentioned topics. In Sect. 27.3, we cover experimental studies in finance focusing on some recent studies, which cast doubt on some of the results and claims of PT and CPT. In Sect. 27.4, we analyze the implication of the experimental findings to the theory of finance. Concluding remarks are given in Sect. 27.5.

27.2 Allias Paradox, PT, CPT, and RDEU: Claims and Implication to the Theory of Finance

27.2.1 Probability Distortions (or Decision Weights)

Most models in economics and finance assume expected utility maximization. Probably the most famous example contradicting the expected utility paradigm is provided by Allias, and is known as the Allias paradox (1953). Table 1 provides two choices in both part I and part II. In part I most subjects would typically choose *A*, while in part II most of them choose *D*. Such choices constitute a contradiction to the classic EU paradigm because from the choice in part I we can conclude that:

Table 1 Allias paradox.
All outcomes are in million \$

Part I			
	A		B
Outcome	Probability	Outcome	Probability
1	1	0	0.01
		1	0.89
		5	0.10
Part II			
	C		D
Outcome	Probability	Outcome	Probability
0	0.89	0	0.90
1	0.11	5	0.10

$$u(1) > 0.01u(0) + 0.89u(1) + 0.10u(5)$$

This inequity can be rewritten as

$$0.11u(1) > 0.01u(0) + 0.10u(5), \quad (1)$$

and the choices in part II implies that

$$0.89u(0) + 0.11u(1) < 0.9u(0) + 0.10u(5)$$

The last inequality can be rewritten also as

$$0.11u(1) < 0.01u(0) + 0.10u(5) \quad (2)$$

As Eqs.(1) and (2) contradict each other for any preference u , we have an inconsistency in the choices in part I and II. How can we explain this result? Does it mean that the EU paradigm is completely wrong? And if the answer is positive, do we have a better substitute to the EU paradigm?

The preference of D over C is not surprising. However, the preference of A over B in Part I seems to induce the paradox. The choice of A is well-known as the “certainty effect,” (see Kahneman and Tversky 1979), i.e. the “one bird in the hand is worth more than two in the bush” effect. The explanation for the contradiction in Eqs.(1) and (2) is due to the “certainty effect,” or alternatively, due to probability distortion in the case where probabilities are smaller than 1. Indeed, experimental psychologists find that subjects tend to subjectively distort probabilities in their decision making. To be more specific, one makes a decision using a weight $w(p)$ rather than the objective probability p . In our specific case, $w(0.01) > 0.01$ – hence the attractiveness of B relative to A decreases, which explains the choice of A in this case. However, in such a case, the classical von Neuman–Morgenstern expected utility is rejected once decision weight $w(p)$ is employed rather than objective probability p .

Probability distortions or decision weights is a subject of many experimental studies conducted mainly by psychologists. Probably the earliest experiments

showing that subjects distort probabilities were conducted by Preston and Baratta (1948) and Edwards (1955, 1962). However, the publication of Prospect Theory (PT) by Kahneman and Tversky in 1979 in *Econometrica* has exposed this issue widely to economists, and hence has strongly influenced research in economics and finance. Though decision weights is an old notion, it is still currently occupying researchers (see for example, Prelec 1998).

In their original paper, Kahneman and Tversky argue that probability p is changed to decision weight $w(p)$ in some systematic manner. However, probability distortion as suggested by Kahneman and Tversky (1979) as well as in the previous studies mentioned above may violate First degree Stochastic Dominance (FSD) or the monotonicity axiom, a property that most economists and psychologists alike are not willing to give up, because violation of FSD essentially means preferring less over more money. Before we illustrate this property, let us first define FSD.

FSD Let F and G be the cumulative distributions of the returns on two uncertain prospects. Then F dominates G by FSD if $F(x) \leq G(x)$ for all x , and there is at least one strict inequity. Moreover,

$$F(x) \leq G(x) \text{ for all } x \Leftrightarrow E_F u(x) \geq E_G u(x) \\ \text{for all utility function } u \in U_1 \text{ where}$$

U_1 is the set of all nondecreasing utility functions ($u' \geq 0$) (see Hanoch and Levy 1969; Hadar and Russell 1969). For a survey and more details, see Levy 1992, 1998.

Let us illustrate with an example why the decision weights framework of PT may lead to a violation of FSD.

Example Consider two prospects x and y . Suppose that x gets the values 3 and 4 with equal probability, and y gets the value 4 with certainty. It is obvious that y dominates x by FSD. Yet, with possible decision weights $w(1/2) = 3/4$ and $w(1) = 1$, we may find a legitimate preference showing a higher expected value for x , i.e. x is preferred to y despite the fact that y dominates x by FSD. For example, for the function $u(x) = x$ (the same is true for many other utility functions), we have

$$EU(x) = \left(\frac{3}{4}\right)3 + \left(\frac{3}{4}\right)4 = \frac{21}{4} = 5\frac{1}{4} > EU(y) \\ = 1 \times 4 = 4$$

Thus, the FSD inferior prospect is selected, which is an undesired result.

Fishburn (1978) shows that this distortion of probability may contradict FSD, or the monotonicity property, which is considered as a fatal flaw of such a probability distortion framework (see also Machina 1994, p. 97). Quiggin (1982) offers a remedy to this problem. He suggests that the probability distortion should be done as a function of the cumulative distribution rather than as a function of the individual probabilities (for more studies along this line, see also Wakker et al. 1994; Yaari 1987; Machina 1994).

According to Quiggin, a given probability p may be distorted in different ways depending on the ranking of the outcome it corresponds to. Thus, the probability $p = 1/4$ may be distorted to different values $w_i(p)$, depending on the rank of the i th outcome. For example, take the following prospect:

$$x = 1, 2, 3, 4$$

$$P(x) = 1/4 \ 1/4 \ 1/4 \ 1/4$$

Then $w(1/4)$ corresponding to $x = 1$ may be larger than $1/4$ and $w(1/4)$ corresponding to $x = 2$ may be smaller than $1/4$ (the opposite relationship is also possible). Thus, the probability distortion is not only a function of the probability p_i but also on the rank of the corresponding outcome, hence the name Rank-Dependent Expected Utility (RDEU). This is in sharp contrast to the decision weights suggested by Kahneman and Tversky in 1979, because by the original PT, $w(1/4)$ is the same for all values and does not depend on the rank of the outcome. Realizing the possible FSD violation, Quiggin (1982) suggests a modification to PT where a transformation of the cumulative distribution is employed. This idea is the basis for Cumulative Prospect Theory (CPT). By this model, the decision weight is also a function of the rank of the outcome. However, unlike Quiggin, Tversky and Kahneman distinguish between negative and positive outcomes. To be more specific, in the CPT framework, the decision weights are given as follows. Consider a prospect $(x_1, p_1; \dots; x_n, p_n)$, where p_i denote the objective probabilities and x_i denote the outcomes. Assume, without loss of generality, that $x_1 \leq \dots \leq x_k \leq 0 \leq x_{k+1} \leq \dots \leq x_n$. The decision weights, which are employed in CPT are given by

$$\begin{aligned} \pi_1 &= w^-(p_1), \pi_n = w^-(p_n), \\ \pi_i &= w^-(p_1 + \dots + p_i) - w^-(p_1 + \dots + p_{i-1}) \\ &\quad \text{for } 2 \leq i \leq k, \\ \pi_i &= w^+(p_i + \dots + p_n) - w^+(p_{i+1} + \dots + p_n) \\ &\quad \text{for } k+1 \leq i \leq n-1, \end{aligned}$$

where w^- and w^+ are weighting functions, which Tversky and Kahneman (1992) experimentally estimate by the functions,

$$\begin{aligned} w^+(x) &= \frac{x^\gamma}{(x^\gamma + (1-x)^\gamma)^{1/\gamma}} \text{ and} \\ w^-(x) &= \frac{x^\delta}{(x^\delta + (1-x)^\delta)^{1/\delta}} \end{aligned} \quad (3)$$

Given these formulas, Tversky and Kahneman find the following estimates: $\hat{\gamma} = 0.61$ and $\hat{\delta} = 0.69$ (see Tversky and Kahneman 1992, pp. 309–312). It can be easily shown that for $\gamma < 1$ and $\delta < 1$, the weighting functions have a reverse

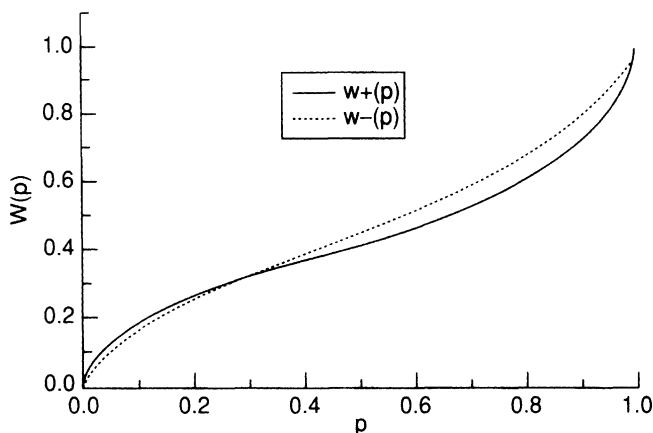


Fig. 1 CPT decision weights

S-shape, implying the overweighing of small probabilities. The probability distortion as suggested by Kahneman and Tversky is illustrated in Fig. 1.

Several researchers argue that in cases of equally likely outcomes, which we call here “uniform” probability distribution, probabilities are not distorted. Quiggin (1982), who was the first one to propose that cumulative probabilities are distorted rather than the raw individual probabilities, argues that for “two equally likely” outcomes ($p = 0.50$) there will be no distortion of probabilities. This argument contradicts Eq. (3), which suggests a distortion even in this case. Though Quiggin does not extend his argument beyond 50:50 bet (actually by his method any other uniform bet, e.g. with 3 or more equally likely outcomes, is distorted) we hypothesized that the probability of a uniform bet (with a $1/n$ probability for each of the n outcomes) should not be distorted as long as the outcomes are not extreme. This is also the result of Viscusi’s (1989) “Prospective Reference Theory” with a symmetric reference point, for which he finds experimental support. However, not all authors agree with the fact that uniform probability distributions are undistorted. Nevertheless, recall that if probabilities are distorted even with a uniform distribution, it has a devastating impact on all reported empirical studies in finance and economics (see below).

The RDEU of Quiggin transforms probabilities in the following manner. Instead of comparing the cumulative distributions F and G , the subjects compare the distributions F^* and G^* where $F^* = T(F)$ and $G^* = T(G)$, where T is the distortion function with $T' > 0$. It can be easily shown that using CPT or RDEU decision weights does not violate FSD. Namely,

$$F^* \leq G^* \Leftrightarrow T(F^*) \leq T(G^*) \quad (4)$$

(See Levy and Wiener 1998. For a survey of SD rules, PT, and the impact of decision weights on choices, see Levy 1998).

In PT and CPT frameworks, probabilities are also distorted in the uniform case. However, the advantage of PT over CPT is that with PT all probabilities with the same size, e.g. $p_i = 1/4$ are distorted in an identical way, hence the choices in a uniform bet are not affected by the probability distortion as suggested by PT. The advantage of CPT over PT is that FSD is not violated. Recalling that CPT decision weights is a technical method which was invented to avoid FSD violations, and that FSD violations do occur experimentally (see Birnbaum 1997) leads one to question the benefit of introducing CPT decision weights.

27.2.2 Change of Wealth Rather than Total Wealth

Expected utility is defined on total wealth, i.e. $u(w + x)$ where w is the initial wealth and x is the change of wealth. Experimental studies reveal that subjects make decisions based on change of wealth, i.e. $u(x)$, rather than $u(w + x)$. It is interesting to note that though the change of wealth argument has been shown experimentally by Kahneman and Tversky, this idea appeared in the literature as early as 1952. Markowitz (1952b) claims that investors make decisions based on change of wealth rather than total wealth. It is easy to construct an example showing that

$$Eu(w + x) > Eu(w + y) \text{ and } Eu(x) < Eu(y)$$

when x and y are the returns on two risky projects. As we shall see later on in this paper, ignoring the initial wealth may indeed affect the choice of the “optimum” portfolio from the efficient set. However, it does not affect the division of the feasible set of portfolios to the efficient and inefficient sets.

27.2.3 Integration of Cash Flows

Expected utility maximization and portfolio selection advocate that one should select a portfolio of assets that maximizes expected utility and one should not consider each asset in isolation. Therefore, correlations should play an important role in portfolio selection. Tversky and Kahneman experimentally find that this is not the case, hence conclude that subjects have difficulties in integrating cash flows from various sources. Let us illustrate this idea with the experiment conducted by Tversky and Kahneman in 1981 with the following two tasks.

Task I: Imagine that you face a pair of concurrent decisions. First, examine both decisions, then indicate the option you prefer.

Decision 1: Choose between A and B given below:

(A) A sure gain of \$2400.

(B) 25% chance to gain \$10,000 and 75% chance to gain nothing.

Decision 2: Choose between C and D given below:

(C) A sure loss of \$7500.

(D) 75% chance to lose \$10,000 and 25% to lose nothing.

A large majority of people choose A in decision 1 and D in decision 2.

Task II: Choose between E and F given below:

(E) 25% chance to win \$2400 and 75% chance to lose \$7600

(F) 25% chance to win \$2500 and 75% chance to lose \$7500.

In Task II, everybody correctly preferred option F over option E . Indeed, F dominates E by FSD. Note that if you return to Task I, however, you get that the inferior option E in Task II is obtained by choosing A and D . The dominating option F in Task II is obtained by combining the two options that most people reject in Task I (i.e. $F = B + C$). Thus, a fully rational decision maker who knows to integrate cash flows from various sources should incorporate the combined decisions, and realize that the combined cash flows of $B + C$ dominate those of $A + D$ in Task I.

From this and other examples, Tversky and Kahneman conclude that investors consider decision problems one at a time instead of adopting a broader frame. Such a procedure induces a reduction in expected utility because the investors miss an opportunity to diversify, hedge, or self-insure. The “narrow framing” of investors arises from the common practice of maintaining multiple “mental accounts.” Thus, the main finding is that the subjects – at least those who participated in the study – are limited in their capability to integrate cash flows from various sources even in a relatively simple case let alone in more complicated cash flows from many sources. If these findings are relevant, not only to subjects in an experiment but also to the investors in practice, this is a severe blow to diversification theory of Markowitz (1952a, 1959, 1987) and Tobin (1958).

27.2.4 Risk Seeking Segment of Preferences

Most models in economics and finance assume risk aversion, i.e. a preference u with $u' > 0$ and $u'' < 0$ (see for example Arrow 1965, 1971; Pratt 1964). However, as early as 1948, Friedman and Savage, based on observed peoples' behavior, suggested a risk-seeking segment of the preference. Markowitz (1952b) modified this preference and suggested another function, which also contains a risk-seeking segment. Both of these studies rely on positive economics arguments. Kahneman and Tversky, on the other hand, base their argument on experimental findings (see also Swalm 1966).

Figure 2 provides the main utility functions advocated in the literature. Figure 2a depicts the classical utility function which is concave everywhere, in accordance with the notion of decreasing marginal utility. Such a function implies risk aversion, meaning that individuals would never accept any fair bet (let alone unfair bets). Friedman and Savage (1948) claim that the fact that investors buy insurance, lottery

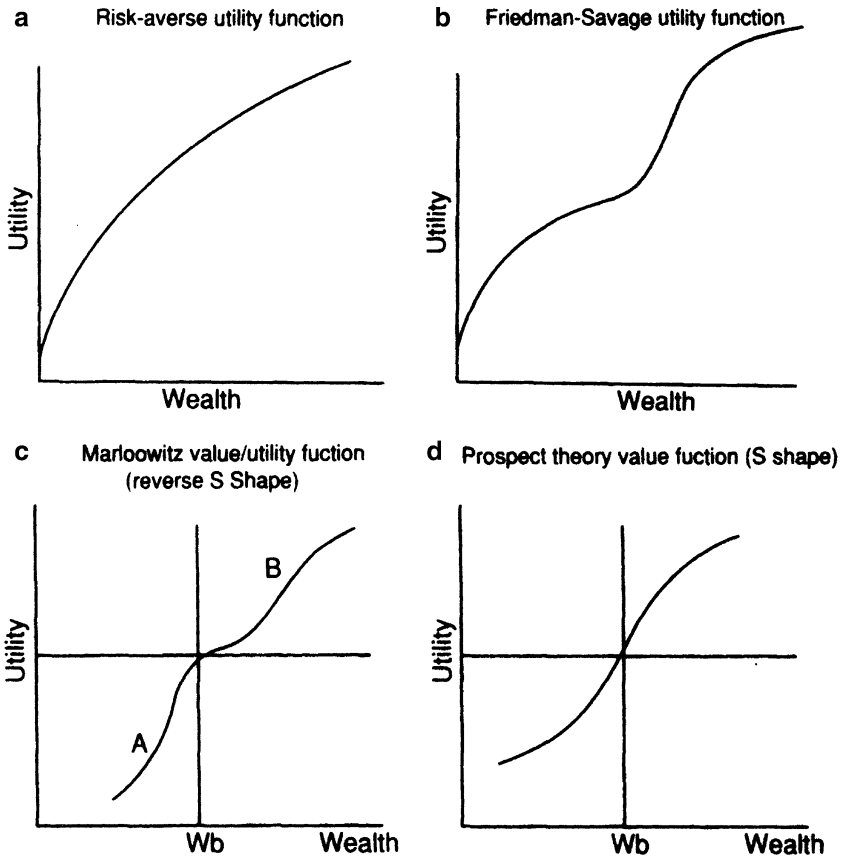


Fig. 2 Alternative shapes of the utility/value function

tickets, and both insurance and lottery tickets simultaneously, plus the fact that most lotteries have more than one big prize, imply that the utility function must have two concave regions with a convex region in between, as represented in Fig. 2b.

Markowitz (1952b) points out several severe problems with the Friedman and Savage utility function. However, he shows that the problems are solved if the first inflection point of the Friedman and Savage utility function is exactly at the individual's current wealth. Thus, Markowitz introduces the idea that decisions are based on "change" in wealth. Hence, Markowitz's utility function can be also considered as a "value function" (as later suggested by Kahneman and Tversky in 1979). By analyzing several hypothetical gambles, Markowitz suggests that individuals are risk-averse for losses and risk-seeking for gains, as long as the possible outcomes are not very extreme. For extreme outcomes, Markowitz argues that individuals become risk-averse for gains, but risk-seeking for losses. Thus, Markowitz suggests a utility function, which is characterized by three inflection points, as

shown in Fig. 2c. Notice that the central part of this function (the range between Points A and B in Fig. 2c) has a reversed S-shape.

Based on their experimental results, with bets which are either negative or positive, Kahneman and Tversky (1979) and Tversky and Kahneman (1992) claim that the value function is concave for gains and convex for losses, yielding an S-shaped function, as shown in Fig. 2d.

27.3 Experimental Studies in Finance

Experimental studies in finance lagged behind experimental studies in economics. Yet, the whole November/December 1999 issue of the *Financial Analyst Journal* is devoted to behavioral finance and discusses issues such as arbitrage, overconfidence, momentum strategies, market efficiency in an irrational world, and equity mispricing. In this section we discuss a few experimental studies in finance.

27.3.1 Portfolio Diversification and Random Walk

In the last three decades, there has been a growing interest of economists in experimental economics. The Nobel prize committee recognized this important field by awarding the Nobel Prize in 2002 to Vernon Smith and Daniel Kahneman. On the importance of experimental research in economics, Vernon Smith asserts: “It is important to economic science for theorists to be less own-literature oriented, to take seriously the data and disciplinary function of laboratory experiments, and even to take seriously their own theories as potential generators of testable hypotheses” (See Smith 1982, p. 924; Plott 1979; Smith 1976, 1982; Wilde 1980).

While laboratory experiments are widely used in economics research, finance research is well behind in this respect. Probably, the first serious experiment in finance was done by Gordon et al. (1972), who studied portfolio choices experimentally. They indicate that to study the investors’ preference, there is an advantage to the experimental method over the empirical method simply because it is difficult, if not impossible, to obtain empirically the relevant data (For a similar argument, see Elton and Gruber 1984). The first experimental studies in finance in diversification and portfolio choices focused on the allocation of money between the riskless asset and one risky asset as a function of various levels of wealth (see Gordon et al. 1972; Funk et al. 1979; Rapoport 1984).

Kroll et al. (1988a) study the choice between risky assets whose returns are normally distributed, where the riskless asset (borrowing and lending) is allowed. The subjects were undergraduate students who did not study finance or investment courses. The main findings of this experiment are:

- (1) The subjects selected a relatively high percentage of mean-variance “inefficient” portfolios.
- (2) The errors involved do not decrease with practice.

- (3) The subjects requested a lot of useless information, i.e. they asked for historical rates of returns when the parameters were known and the returns were selected randomly (information given to the subjects). Thus, the subjects presumably believed that there are some patterns in rates of return, though such patterns do not exist.

Odean (1998) in his analysis of many individual transactions reports that there is a tendency of investors to hold losing investments too long and to sell winning investments too soon (a phenomenon known as the disposition effect). This result is consistent with the results of Kroll et al. (see (3) above). He finds that when individual investors sold a stock and quickly bought another, the stock they sold outperformed on average the stock they bought by 3.4% points in the first year. This costly overtrading may be explained by the fact that investors perceive patterns where none exist or do not want to admit their errors in selection of their investments. Perceiving patterns when they do not exist is exactly as reported by KL&R.

In a subsequent paper, Kroll et al. (1988b) experimentally test the Separation Theorem and the Capital Asset Pricing Model (CAPM). In this experiment the 42 subjects were undergraduate students who took a course in statistics. They had to select portfolios from three available risky assets and a riskless asset. This experiment reveals some negative and some positive results. The results are summarized as follows:

- (1) As predicted by the M-V rule, the subjects generally diversified between the three riskless assets.
- (2) A tenfold increase in the reward to the subject significantly improved the subjects' performance. This finding casts doubt on the validity of the results of many experiments on decision making under uncertainty which involves a small amount of money.
- (3) Though the subjects were told that rates of return are drawn randomly, as before, they, again, asked for (useless) information. This finding may explain why there are "chartists" and "technical analysts" in the market even if indeed rates of returns are randomly distributed over time. Thus, academicians may continue to claim the "random walk" property of returns and practitioners will continue to find historical patterns and employ technical rules for investment based on these perceived patterns.
- (4) Changing the correlation (from -0.8 to 0.8), unlike what Markowitz's theory advocates, does not change the selected diversification investment proportions.
- (5) The introduction of the riskless asset does not change the degree of homogeneity of the investment behavior. Thus, at least with these 42 subjects the Separation Theorem (and hence the CAPM) does not hold in practice.

In the study of KL&R, the subjects were undergraduate students with no background in finance and they could not lose money. These are severe drawbacks as the subjects may not represent potential investors in the market. To overcome these drawbacks, Kroll and Levy (K&L) (1992) conducted a similar experiment with the

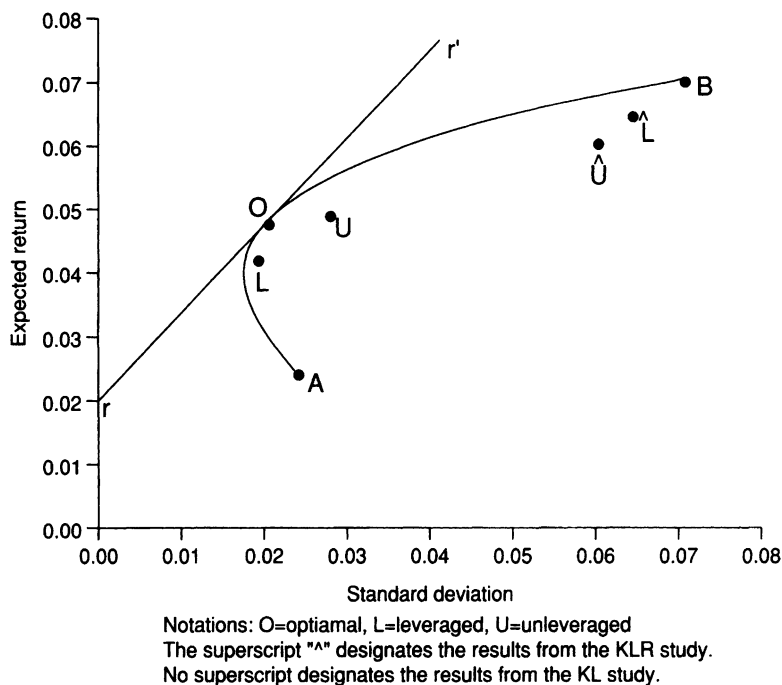


Fig. 3 The efficient frontier and the actual portfolios selected by the subjects

same parameters as in KL&R but this time with second year MBA students and where financial gains and losses were possible. The results improved dramatically in favor of Markowitz's diversification theory. Figure 3 shows the average portfolio with and without leverage selected in the KL&R study and in the K&L study. In the K&L study, the selected portfolios are L and U (for levered and unlevered portfolios) while in KL&R they are $\hat{U}$ and $\hat{L}$. As can be seen U and L are much closer to the optimum solution (in particular, L is much closer to line rr' than $\hat{L}$), indicating that when real money is involved and MBA students are the subjects, much better results are achieved. Also, as predicted by portfolio theory, the subjects, unlike in KL&R study, change the investment proportions when correlation changes.

Finally, the investment proportions selected were similar to those of the optimum mean variance portfolio. Therefore, K&L conclude that the subjects behave *as if* they solve a quadratic programming problem to find the optimum portfolio even though they did not study this tool at the time the experiment was conducted.

27.3.2 The Equity Risk Premium Puzzle

The difference between the observed long-run average rate of return on equity and on bonds cannot be explained by well behaved risk averse utility function; hence

the term equity risk premium puzzle. Benartzi and Thaler (1995) suggest the loss aversion preference as suggested by PT S-shape function to explain the existing equity risk premium. They show that if investors weight losses 2.5 more heavily than possible gains the observed equity risk premium can be explained. However, Levy and Levy (2002b) have shown that the same conclusion may be drawn with a reverse S-shape utility function as suggested by Markowitz, as long as the segment corresponding to $x < 0$ is steeper than the segment corresponding to $x > 0$.

Levy and Levy (2002c) (L&L) analyze the effect of PT and CPT decision weights on Arrow (1965) and Pratt (1964) risk premium. They show that a positive risk premium may be induced by decision weights $w(p)$ rather than probabilities p even in the absence of risk aversion. In their experiment a large proportion of the choices contradicts risk aversion but this does not contradict the existence of a positive equity risk premium. Thus, one does not need loss aversion to explain Arrow's risk premium because it can be induced by the use of decision weights. Unlike the case of Arrow's risk premium, with Pratt's risk aversion measure or with historical data, which is composed of more than two values, the risk premium may increase or decrease due to the use of decision weights. To sum up, the equity risk premium puzzle can be explained either by loss aversion, which is consistent both with an S-shape function and a reverse S-shape function, or by decision weights, even in the absence of loss aversion.

27.3.3 The Shape of Preference

Risk aversion and a positive risk premium are two important features of most economic and finance models of assets pricing and decision making under uncertainty. Are people risk averse? As shown in Fig. 1, Friedman and Savage, Markowitz, and K&T claim that this is not the case. So what can we say about preference? In a series of experiments with and without financial rewards, Levy and Levy (2002a, b) have shown that a major portion of the choices contradict risk-aversion. L&L conducted several experiments with 328 subjects. To test whether the subjects understood the questionnaire and did not fill it out randomly just to "get it over with," they first tested FSD which is appropriate for risk-seekers and risk averters alike. They found that 95% of the choices conform with FSD (i.e. with the monotonicity axiom), which validates the reliability of their results. They find that in Experiment 1 *at least 54%* of the choices contradict risk aversion, in Experiment 2 *at least 33%* of the subjects contradict risk aversion and in Experiment 3 *at least 42%* of the choices contradict risk aversion. It is interesting to note that the subjects in these three experiments were business school students, faculty, Ph.D students and practitioners (financial analysts and funds managers).

In these three experiments, L&L also tested the effect of the subjects characteristics, the size of the outcomes as well as the framing of the bet. The results are very similar across all these factors with the exception that Ph.D. students and

faculty members choose more consistently with risk aversion (71–78% correct second degree stochastic dominance (SSD) choices). But there may be a bias here because these subjects are more familiar with SSD rules, and it is possible that they mathematically applied it in their choices. However, even with these sophisticated subjects at least 22–29% of them selected inconsistently with risk aversion, implying trouble for theoretical models, which rely on risk aversion.

The fact that 33–54% of the subjects behave “as if” they are not risk averse, implies that they choose “as if” the utility function is not concave in the whole range. For example, K&T, Friedman and Savage and Markowitz utility functions are consistent with L&L findings. Note that L&L findings do not imply risk seeking in the whole range, but rather no risk-aversion in the entire range. Therefore, their finding does not contradict the possibility that with actual equity distribution of rates of return corresponding to the US market the risk premium may even increase due to decision weights.

Rejecting risk aversion experimentally is repeated in many experiments (see for example L&L, 2001, 2002a). Hence, the remainder contrasts K&T S-shape function and Markowitz’s reverse S-shape function. Employing prospect stochastic dominance (PSD) and Markowitz’s stochastic dominance (MSD) L&L contrast these two utility functions. Let us first present these two investment criteria:

27.3.3.1 Prospect Stochastic Dominance (PSD)

Let U_s be the set of all S-shape preferences with $u' > 0$ for all $x \geq 0$ and $u'' > 0$ for $x < 0$ and $u' < 0$ for $x > 0$. Then

$$\int_y^x [G(t) - F(t)] dt \geq 0 \text{ for all } x > 0, y < 0 \quad (5)$$

$$\Leftrightarrow E_F u(x) \geq E_G u(x) \text{ for all } u \in U_s.$$

27.3.3.2 Markowitz Stochastic Dominance (MSD)

Let U_M be the set of all reverse S-shape preferences with $u' > 0$ for all $x \geq 0$ and $u' < 0$ for $x < 0$ and $u' > 0$ for $x > 0$. Define F and G as above. Then F dominates G for all reverse S-shaped value functions, $u \in U_M$, if and only if

$$\int_{-\infty}^y [G(t) - F(t)] dt \geq 0 \text{ for all } y$$

$$\leq 0 \text{ and } \int_x^{\infty} [G(t) - F(t)] dt$$

$$\geq 0 \text{ for all } x \geq 0 \quad (6)$$

(with at least one strict inequality). And Eq. (5) holds iff $E_F u(x) > E_G u(x)$ for all $u \in U_M$. We call this dominance relation MSD—Markowitz Stochastic Dominance.

Table 2 presents the four Tasks while Table 3 presents the results of the experimental study of Levy and Levy (2002a). Note that in Task I G dominates F by PSD, but F dominates G by MSD. As can be seen from Table 3, in Task I, 71% of the subjects choose F despite the fact that G dominates F by PSD. Thus, *at least* 71% of the choices are in contradiction to PSD, and supporting MSD, i.e. a reverse S-shape preference as suggested by Markowitz. Note also that 82–96% of the choices (see Tasks II and III) are consistent with FSD. Once again, by the results of Task IV we see that about 50% of the choices reject the assumption of risk aversion (SSD). Table 5 taken from Levy

Table 2 The choices presented to the subjects

Suppose that you decided to invest \$10,000 either in stock F or in stock G. Which stock would you choose, F, or G, when it is given that the *dollar gain or loss* 1 month from now will be as follows

TASK I				
	F Gain or loss	Probability	G Gain or loss	Probability
	−3000	1/2	−6000	1/4
	4500	1/2	3000	3/4

Please write F or G

TASK II				
Which would you prefer, F or G, if the dollar gain or loss 1 month from now will be as follows				
	F Gain or loss	Probability	G Gain or loss	Probability
	−500	1/3	−500	1/2
	+2500	2/3	2500	1/2

Please write F or G

TASK III				
Which would you prefer, F or G, if the dollar gain or loss 1 month from now will be as follows				
	F Gain or loss	G Probability	Gain or loss	Probability
	+500	3/10	−500	1/10
	+2000	3/10	0	1/10
	+5000	4/10	+500	1/10
			+1000	2/10
			+2000	1/10
			+5000	4/10

Please write F or G

TASK IV				
Which would you prefer, F or G, if the dollar gain or loss 1 month from now will be as follows				
	F Gain or loss	G Probability	Gain or loss	Probability
	−500	1/4	0	1/2
	+500	1/4		
	+1000	1/4	+1500	1/2
	+2000	1/4		

Please write F or G

Source: Levy and Levy (2002a)

Table 3 The results of the experiment^a

Task	F	G	Indifferent	Total
I(GϕPSDF,FϕMSDG)	71	27TT	2	100
II(FϕFSDG)	96	4	0	100
III(FϕFSDG)	82	18	0	100
IV(GϕSSDF)	47	51	2	100

Source: Levy and Levy (2002a)
Number of subjects: 260
^aNumbers in the tables are in percent, rounded to the nearest integer.
The notations ϕFSD, ϕSSD, and ϕMSD indicate dominance by FSD, SSD, PSD, and MSD, respectively

and Levy (2002b) reveals once again the results of another experiment showing that at least 62% of the choices contradict the S-shape preference of PT.

Wakker (2003) in his comment on Levy and Levy’s (2002a) paper claims that the dominance by PSD (or by MSD) also depends on probability weights. Generally, his claim is valid. However, if a uniform distribution ($p_i = 1/4$ for all observations) is considered, his criticism is valid only if indeed probabilities are distorted in such a case. Are probabilities distorted in such a case? And if the answer is positive, can we blindly use the distortion formula suggested by T&K? There is evidence that in the case of uniform distributions probabilities are not distorted, or are distorted as recommended by PT but not by CPT (hence do not affect choices). Thus, Wakker’s claim is invalid. Let us elaborate. In Viscusi’s (1989) *Prospective Reference Theory* there is also *no* probability distortion in the symmetric case. Also, in the original PT framework (Kahneman and Tversky 1979), in which the probabilities are transformed directly, the choice among prospects is unaffected by subjective probability distortion in the case of equally likely outcomes. Thus, as the study reported in Table 4 was conducted with uniform probabilities and moderate outcomes, it is safe to ignore the effects of subjective probability distortion in this case. Wakker (2003) argues that by CPT probabilities are distorted even in the bets given in Table 5, hence the conclusion against the S-shape function by Levy and Levy is invalid. If one uses the distortion formula (see Eq. 3) of T&K also in the uniform case, Wakker is correct. However, recall that the formula of T&K is based on aggregate data of nonsymmetrical probability distributions and with no financial reward or penalty. So why should one think it is appropriate to apply it to the uniform probability case? Moreover, as we see in Sect. 27.2.1, Eq. (3) suggests decision weights which are hard to accept if indistinguishability is employed in all cases.

Yet, even if one adheres to T&K distortion weights formula, even in the equally likely outcomes case, the S-shape preference is rejected and Wakker is wrong in his criticism. Indeed, Levy and Levy (2002b) conduct a direct confrontation of PSD and MSD where probability distortion is taken into account exactly as suggested by K&T’s CPT and exactly as done by Wakker (2003). Table 6 presents the two choices *F* and *G*, the objective probabilities as well as the decision weights as recommended by CPT [see Eq. 3].

Table 4 The choices presented to the subjects

Suppose that you decided to invest \$10,000 either in stock F or in stock G. Which stock would you choose, F, or G, when it is given that the *dollar gain or loss* 1 month from now will be as follows

	F		G	
	Gain or loss	Probability	Gain or loss	Probability
	−1600	1/4	−1000	1/4
	−200	1/4	−800	1/4
	1200	1/4	800	1/4
	1600	1/4	2000	1/4
Please write F or G				

Table 5 The results of experiment 2^a

	F	G	Indifferent	Total
($F\phi_{\text{PSD}}G, G\phi_{\text{MSD}}F$)	38%	62%	0%	100%

Source: Levy and Levy (2002b)

Number of subjects: 84

^aNumbers in the tables are in percent, rounded to the nearest integer. The notations ϕ_{PSD} , and ϕ_{MSD} indicate dominance by PSD, and MSD, respectively

Table 6 G dominates F by PSD even with CPT decision weights (task II of experiment 3 in Levy and Levy 2002b)

F			G		
Gain or loss	Probability	CPT decision weights	Gain or loss	Probability	CPT decision weights
−875	0.5	0.454	−1000	0.4	0.392
2025	0.5	0.421	1800	0.6	0.474

Source: Levy and Levy (2002b)

Note that with the data of Table 6, G dominates F by PSD with objective as well as subjective probabilities. Yet 50% of the subjects selected F . This implies that *at least* 50% (it may be much larger than 50% but this cannot be proven) of the subjects' choices do not conform with an S-shape preference, rejecting this important element of PT and CPT (see Table 6).

To sum up, with objective probabilities the S-shape preference is rejected. With PT the S-shape is also rejected by the decision weights $w(1/4)$ which is identical for all outcomes (For a proof, see Levy and Levy 2002b). Table 6 reveals that the S-shape function is rejected also when decision weights are taken into account exactly as recommended by CPT's formula.

Thus, more than 50% of the choices contradicts risk aversion and more than 50% of the choices contradicts the S-shape function – the preference advocated by PT and CPT. The experiments' results yield most support Markowitz's reverse-

Shape preference. From this above analysis we can conclude that investors are characterized by a variety of preferences and that there is no one dominating preference.

27.3.4 Asset Allocation and the Investment Horizon

Benartzi and Thaler (1999) (B&T) present subjects with a gamble reflecting a possible loss. The subjects could choose to gamble or not in an experiment which contains N repetitions. The subjects were reluctant to take the gamble. However, where the multi-period distributions of outcome induced by the N repetitions was presented to them, more subjects were willing to take the gamble. This shows that the subjects either have difficulties to integrate cash flows from various trials or they have “narrow framing.” These results also have strong implications to asset allocation and the investment horizon. B&T found the subjects willing to invest a substantially higher proportion of their retirement funds in stocks (risky assets) once they were shown the distributions of the long-run return relative to the investment proportion when the distribution of return is not shown to them. The results of B&T shed light on the debate between practitioners and academicians regarding the relationship between the portfolio composition and the investment horizon. While Samuelson (1994) and others correctly claim that for myopic (power) utility functions, the investment horizon should not have any effect on asset allocation, practitioners claim that the longer the horizon, the higher the proportion of assets that should be allocated to stocks. The results of B&T support the practitioners’ view provided that the subjects observe the multi-period distribution, i.e. overcoming the “narrow framing” effect. Ruling out irrationality or other possible biases, this finding means that the subjects in B&T’s experiment do not have a myopic utility function. It is interesting that Leshno and Levy (2002) have shown that as the number of period N increases, stocks “almost” dominate bonds by FSD, when “almost” means for almost all preferences, not including the myopic function. Thus, this theoretical result is consistent with B&T’s experimental results.

27.3.5 Diversification: The 1/n Rule

Let us open this section by the following old assertion:

Man should always divide his wealth into three parts: one third in land, one third in commerce and one third retained in his own hands.

Babylonian Talmud

Two interesting conclusions can be drawn from this 1500-year-old recommendation, which is probably the first diversification recommendation. The first conclusion is consistent with what Markowitz recommended and formalized about 50 years ago:

diversification pays. The second conclusion is in contrast to Markowitz's recommendation: invest 1/3 in each asset and ignore the optimum diversification strategy, which is a function of variances, correlations, and means.

It is interesting that Benartzi and Thaler (2001) experimentally find that this is exactly what investors do. Presented with n assets (e.g. mutual funds) the subject is inclined to invest $1/n$ in each fund. This is true regardless of the content of funds. If one fund is risky (stocks) or riskless, this does not change the $1/n$ choice, implicitly implying that the Talmud's recommendation is intact as correlations, means and variances are ignored. From this we learn that investors believe that "a little diversification goes a long way," but mistakenly ignore the optimal precise diversification strategy.

27.3.6 The CAPM: Experimental Study

One of the cornerstones of financial theory is asset pricing, as predicted by the CAPM. The problem with testing the CAPM empirically is that the *ex ante* parameters may change over time. However, while the CAPM cannot be tested *empirically* with *ex-ante* parameters, it can be tested *experimentally* with *ex-ante* parameters. The subjects can provide buy-sell orders and determine collectively equilibrium prices of risky assets when the future cash flow (random variables) corresponding to the various assets are given. Levy (1997) conducted such an experiment with potential financial loss and reward to the subjects. Thus, like in Lintner's (1965) approach for given distributions of end-of-period returns, the subjects collectively determine, exactly as in an actual market, the current market values P_{io} . Therefore, the means μ_i variances σ_i^2 and correlations, R_{ij} are determined simultaneously by the subjects. Having these parameters one can test the CAPM with *ex-ante* parameters. Lintner (1965) found that subjects typically diversify in only 3–4 assets (out of the 20 available risky assets), yet the CAPM, or the $\mu - \beta$ linear relationship was as predicted by the CAPM with an R^2 of about 75%. Thus, Levy found a strong support to the CAPM with *ex-ante* parameters.

27.4 Implication of the Experimental Findings to Finance

27.4.1 Arbitrage Models

Let us first analyze the arbitrage-based models like Modigliani and Miller (M&M) (1958), Black and Scholes (1973) model and the APT model of Ross (1976). Let us first illustrate M&M capital structure with no taxes. Denoting by V_U and V_L the value of the unlevered and levered firms, respectively, M&M claim that $V_U = V_L$ further, if $V_U \neq V_L$, one can create an arbitrage position such that the investor who holds return $\tilde{y}$ will get after the arbitrage $\tilde{y} + a$ when $a > 0$. Thus, an FSD position is created and as the two returns are fully correlated, the FSD dominance implies an arbitrage

position. Thus, arbitrage is achieved by selling short the overpriced firm's stock and holding long the underpriced firm's stock. If probability is distorted, this will not affect the results as the investor ends up with the same random variable. If preference is S-shaped, it does not affect the results as FSD position is created, which holds for all $u \in U_1$ and the S-shape functions are included in U_1 . Making decisions based on change of wealth rather than total wealth also does not affect these results (recall that FSD is not affected by initial wealth). However, the fact that investors have difficulties in integrating cash flows from various sources may affect the result. The reason is that before the arbitrage, the investor holds, say, the stock of the levered firm. If $V_L > V_U$ the investor should sell the levered firm, borrow and invest in the unlevered firm. However, the investor should be able to *integrate* the cash flows from these two sources and realize that the levered firm's return is duplicated. According to prospect theory, "mental departments" exists and the subjects may be unable to create this cash flow integration. However, recall that to derive the condition $V_L = V_U$, it is sufficient that *one* investor will be able to integrate cash flows to guarantee this equilibrium condition and not that all investors must conduct this arbitrage transaction ("money machine" argument).

Black and Scholes (B&S) equilibrium option pricing is based on the same no-arbitrage idea. Whenever the call option deviates from B&S equilibrium price economic forces will push it back to the equilibrium price until the arbitrage opportunity disappears. This is very similar to M&M case; hence it is enough that there is one sophisticated investor in the market who can integrate cash flows.

Thus, a "money machine" is created whenever the market price of an option deviates from its equilibrium price. The same argument holds for all theoretical models which are based on an arbitrage argument, e.g., Ross's (1976) arbitrage pricing theory (APT). Thus, for arbitrage models, the integration of cash flows issue may induce a problem to some investors but luckily, in these models, *one* sophisticated investor who knows how to integrate cash flows is sufficient to guarantee the existence of an asset price as implied by these models.

Despite this argument, in a multiperiod setting, when the investment horizon is uncertain, in some cases we do not have a pure arbitrage as the gap in the price of the two assets under consideration (e.g. $V_L > V_U$) may even (irrationally) increase over time (see Thaler 1993). Thus, while in a one-period model where all assets are liquidated at the end of the period, the above argument is valid; this is not necessarily the case in a multi-period setting.

27.4.2 Stochastic Dominance (SD) Rules

It is easy to show that prospect F dominates prospect G in terms of total wealth ($W + x$) if only F dominates G with change in wealth (x). Thus, shifting from *total* wealth to *change* of wealth does not affect the dominance result. The same is true with the Mean-Variance rules. SD rules deal mainly with two distinct options and not with a mix of random variables, hence generally the issues of integration of cash flows does not arise with the application of SD rules (it is relevant, however, to the Mean-Variance rule). If

preference is an S-shaped or reverse S-shape, FSD is intact as it is defined for all $u \in U_1(u' > 0)$. If probabilities are distorted by a distortion function $T(\cdot)$ where $F^* = T(F)$ and $T' > 0$, the FSD relationship is also unaffected by the distortion. Let us turn now to SSD and TSD. If preference is S-shape, SSD or TSD rules which assumes $u' < 0$ are irrelevant. However, even with risk aversion, with probability distortion $T(F)$ ($T' > 0$), the SSD and TSD dominance relationships are affected. Thus, FSD is unaffected by the experimental findings of probability distortion, but SSD and TSD are affected. With probability distortion one should first transform probabilities and only then compare F^* and G^* when $F^* = T(F)$ or $G^* = T(P)$. However, as the k th individual is characterized by probability distortion T_k , each investor has his subjective SD efficient set and therefore the classical two-step portfolio selection (i.e. determining the efficient set in its first step and selecting the optimal portfolio from the efficient set in the second step) is meaningless as there is no one efficient set for all investors. To sum up, FSD efficiency analysis is intact and SSD and TSD are not. If, however, in SSD analysis, it is assumed that $T' > 0$ and $T' < 0$, SSD analysis also remains intact (see Levy and Wiener 1998).

27.4.3 Mean-Variance (M-V) Rule and PT

If the utility function is S-shaped or reverse S-shaped the Mean-Variance rule does not apply and hence it is not an optimal investment decision rule even in the face of normal distributions. Probability distortion is even more devastating to the M-V efficiency analysis. The whole idea of the M-V efficient set is that all investors face the same efficient set which depends on mean, variances and correlations. Now if the k th investor distorts F_i to $T_k(F_i)$ when F_i is the cumulative distribution of the i th asset, then we have K subjective efficient sets (composed of the individual assets) and the idea of M-V efficiency analysis breaks down. Nevertheless, as we shall see below, the M-V efficiency analysis surprisingly is intact in the presence of mutual funds, or if the probability distortion is done on portfolios but not on each individual asset.

27.4.4 Portfolios and Mutual Funds: Markowitz's M-V Rule and PT – A Consistency or a Contradiction?

It is interesting that in the same year (1952) Markowitz, published two seminal papers that seem to contradict one another. One of these papers deals with the M-V rule (Markowitz 1952a) and the other with the reverse S-shape function (Markowitz 1952b). The Mean-Variance rule implies implicitly or explicitly risk aversion, while the reverse S-shape function, and for that matter also the S-shape function of PT, imply that risk aversion does not hold globally. Do we have here a contradiction between these two articles of Markowitz?

To analyze this issue we must first recall that the M-V rule is intact in two alternate scenarios: (1) a quadratic utility function; and (2) risk aversion and normal

distributions of return. Obviously, under S-shape or reverse S-shape preference, scenario (1) does not hold. What about scenario (2)? Generally, if one compares two assets X and Y , indeed it is possible that X dominates Y by the M-V rule, but the expected utility of Y is greater than the expected utility of X for a given S-shape function even if X and Y are normally distributed. Hence, for such a comparison of any two arbitrary prospects, the two papers of Markowitz are indeed in contradiction. However, when diversification among all available assets is allowed, Levy and Levy (2004) have shown that the two articles do not contradict each other but conform with each other. Let us be more specific. Figure 4 illustrates the M-V efficient frontier. It is possible that asset a dominates b by M-V rule but not for all S-shape or reverse S-shape functions. However, and that is the important point, for any interior asset like asset b , there is an asset b' on the frontier which dominates b by the M-V rule as well as by EU for all monotonically increasing utility functions, including the S-shape and reverse S-shape functions. To see this claim, recall that under scenario (2) normal distributions are assumed. Regarding assets b and b' we have $\mu(b') > \mu(b)$ and $\sigma(b') = \sigma(b)$; hence portfolio b' , dominates portfolio b by FSD, or $Eu(b') > Eu(b)$ for all utility functions, including the S-shape and reverse S-shape preferences (For FSD in the normal distribution case see Levy 1998). Thus, Markowitz's M-V diversification analysis is intact for all $u \in U_1$ including $u \in U_M$ and $u \in U_{PT}$ as $U_M \subset U_1$ and $U_{PT} \subset U_1$. To sum up, the quadratic utility function does not conform with the experimental findings regarding preferences. Assuming normality and risk aversion is also not justified in light of the experimental findings regarding preferences. However, allowing investors to diversify (with normal distributions), the M-V efficient frontier is the efficient one

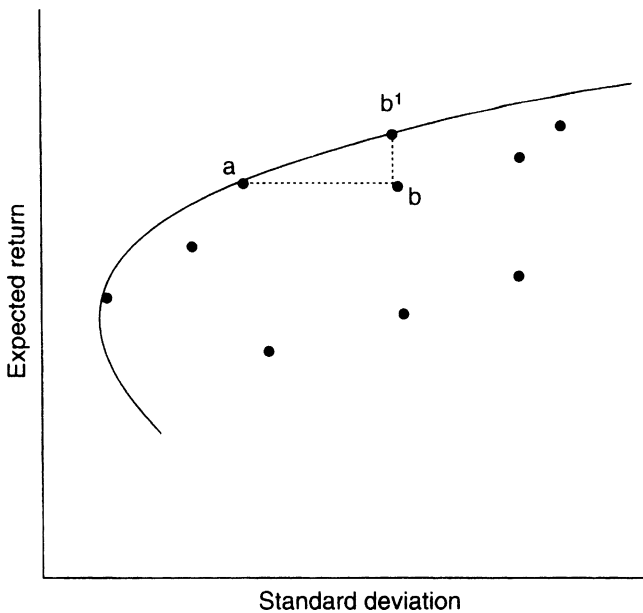


Fig. 4 Mean-variance dominance and FSD dominance

also is EU framework, without the need to assume risk aversion. Also, as the FSD efficient set is invariant to change in wealth instead of total wealth, looking at the change of wealth rather than the total wealth does not change our conclusion. Thus, the two papers of Markowitz are not in contradiction as long as diversification is allowed, an assumption which is well accepted. Thus, while portfolio *a* dominates *b* with risk aversion, such dominance does not exist with other preferences with a risk seeking segment. But a vertical comparison (e.g. *b* and *b'*) allows us to conclude that Markowitz's M-V inefficient set is inefficient for all $u \in U_1$ and not only to risk averse utility functions (see Fig. 4).

So far, we have dealt only with the factors of preferences and change of wealth rather than total wealth. Let us now see how the other findings of PT and CPT affect the M-V analysis. First, if individual investors fail to integrate cash flows from various assets, the M-V efficiency analysis and the Separation Theorem collapse. However, we have mutual funds and in particular indexed funds which carry the integration of cash flows for the investors. If such mutual funds are available – and in practice they are – the M-V analysis is intact also when one counts for the “mental departments” factor. To see this, recall that with a normal distribution mutual fund *b'* dominates mutual fund (or asset) *b* by FSD with objective distributions because $F_{b'}(X) \leq F_b(X)$ for all values *x* (see Fig. 4). As FSD is not affected by CPT probability distortion, also $T(E_{b'}(X)) \leq T(E_b(X))$, hence the M-V efficient set (of mutual funds) is efficient also in CPT framework. Thus, change of wealth (rather than total wealth), risk-seeking segment of preference and probability distortion (as recommended by CPT), do not affect Markowitz's M-V efficiency analysis and the Separation Theorem. Hence, under the realistic assumption that mutual funds exist, the CAPM is surprisingly intact even under the many findings of experimental economics which contradict the CAPM assumptions.

27.4.5 The Empirical Studies and Decision Weights

Though we discuss above the probability distortion in the uniform case, we need to return to this issue here as it has a strong implication regarding empirical studies in finance.

Nowadays, most researchers probably agree that probability distortion takes place at least at the extreme case of low probabilities (probability of winning in a lottery, probability that a fire breaks out, etc.). Also probability distortion may take place when extreme gains or losses are incurred. Also the “certainty effect” is well documented. However, if one takes probability distortion too seriously and adopts it exactly as recommended by T&K (1992) [see Eq. 3 above], paradoxes and absurdities emerge. For example, take the following two prospects:

x:	−\$2000	−\$1000	+\$3000	+\$4000
p(x)	1/4	1/4	1/4	1/4
y:	−\$10 ⁶	−\$1000	+\$10 ¹²	\$10 ²⁴
p(y)	1/4	1/4	1/4	1/4

Employing the distortion of T&K [see Eq. 3 above] for probability distortion implies the following decision weights:

x :	−\$2000	−\$1000	+\$3000	+\$4000
$w(x)$	0.29	0.16	0.13	0.29
y :	−\$10 ⁶	−\$1000	+\$10 ¹²	+\$10 ²⁴
$p(y)$	0.29	0.16	0.13	0.29

Does it make sense? With x , the probability of $x = -2000$ increases from 0.25 to 0.29, and the probability of $x = -1000$ drops from 0.25 to 0.16. And similar is the case with y . Thus, the magnitudes of the outcomes are not important and the probabilities are distorted by the same formula regardless of whether the outcome is -10^6 or only -2000 . This type of probability distortion which is insensitive to the magnitude of the outcomes, has been employed by Wakker (2003). Not everyone agrees with CPT weighting function [see Eq. 3]. Birnbaum and McIntosh (1996), finds that the probability distortion depends on the configuration of the case involved, hence suggests a “configurational weighting model.” Moreover, Birnbaum experimentally shows that in some cases, FSD is violated. Note that FSD was the main reason why CPT and RDEU (RDEU) were suggested as substitutes to PT. Thus, if FSD is indeed violated, CPT is losing ground and PT may be a better description of investors’ behavior. Similarly, Quiggin (1982) suggests that in cases where there are two equally likely outcomes with a 50:50 chance, probability is not distorted, which contradicts Eq. (3). Viscusi (1989) shows that with probability $p_i = 1/n$ with n possible outcome, $w(1/n) = 1/n$, i.e. probability is not distorted. Also, the original PT of K&T implies that in case of an equally likely outcome, the choices are not affected by the decision weights. We emphasize this issue because the issue whether probability is distorted or not in equally likely outcomes has important implications to empirical studies.

In virtually all empirical studies in finance and economics where distribution is estimated, n observations are taken and each observation is assigned an equal probability. For example, this is the case in the calculation of σ^2 , β , etc. If probabilities are also distorted in the uniform case, namely, $w(1/n) \neq 1/n$ as suggested by CPT of T&K, all the results reported in the empirical studies are questionable, including all the numerous empirical studies which have tested the CAPM. The implicit hypothesis in these studies is that in the uniform probability case, with no extreme values, probabilities are not distorted. We cannot prove this, but recall that the distortion formula of T&K is also obtained in a very limited case; an experiment with some specific lotteries when the S-shaped preference and the weighting function are tested simultaneously. Hence, the parametric assumptions concerning both functions are needed. Also, T&K report aggregate rather than individual results. Therefore, the burden of the proof that probabilities are distorted in the uniform case is on the advocates of PT and CPT. Finally, as mentioned above, if one employs PT (1979), decision weights where $w(1/n) = p_o$ for all observations, the decision maker who maximizes EU will not change his decision also in the uniform case (for a proof see, Levy and Levy 2002b). However, with decision weight p_o rather than $1/n$, the empirical results may change, despite the fact that choices are unchanged. This issue has to be further investigated.

27.5 Conclusion

Experimental research is very important as it allows us to control variables and sometimes to study issues that cannot be studied empirically, e.g. testing the CAPM with *ex-ante* parameters. Experimental findings and in particular Prospect Theory and cumulative Prospect Theory (PT and CPT, respectively) contradict expected utility theory (EUT) which, in turn, may have a direct implication to financial and economic decision making theory and to equilibrium models.

Taking the PT and CPT implication to the extreme, we can assert that virtually all models in finance and in particularly all empirical studies results and conclusions are incorrect. However, this conclusion is invalid for two reasons: (a) subjects in the experiments are not sophisticated investors who in practice risk a relatively large amount of their own money; and (b) one cannot conclude from a specific experiment (or from a few of them), conducted with no real money and with unequal probabilities that probabilities are distorted also in the uniform case, i.e. with equally likely outcomes. Therefore, PT and CPT do not have an unambiguous implication regarding the validity of the empirical studies in economics and finance which implicitly assume equally likely outcomes. Let us elaborate.

Drawing conclusions from the “average subject behavior” and assuming that also sophisticated investors behave in a similar way may be misleading. For example, for all arbitrage models it is possible that most subjects and even most investors in practice do not know how to integrate cash flows, but it is sufficient that there are some sophisticated investors who integrate cash flows correctly to obtain the arbitrage models’ equilibrium formulas (e.g. APT, M&M and B&S models). Probability distortion, various preference shapes and change of wealth rather than the total wealth, do not affect these arbitrage models. There are exceptions however. If the gap between the prices of the two assets increases irrationally rather than decreases over time, the arbitrage profit is not guaranteed (see Thaler 1999). To sum up, in a one-period model where the assets are liquidated at the end of the period, the arbitrage models are intact, but this is not necessarily the case in a multiperiod setting with irrational asset pricing.

The capacity of investors to carry cash-flows integration is crucial in particular for M-V portfolio selection. However, in this case we may divide the group of investors into two subgroups. The first group is composed of sophisticated investors who diversify directly; hence presumably do not conduct the common mistakes done by subjects in experiments. This group includes also the financial consultants and advisers who know very well how to take correlations of returns (integration of cash flows) into account. For example, professional advisers recommend that, “You don’t want more than one company in an industry, and you don’t want companies in related industries.”

This advice is quite common and one can document numerous other similar assertions made by practitioners. Therefore, it is obvious that cash-flows integration and correlations are well taken into account by this segment of investors. The other group of investors who may be exposed to all deviations from rationality are composed of the less sophisticated investors. These investors who cannot integrate cash flows from various sources may buy mutual funds managed by professional investors. Thus, if the M-V

efficient set contains these mutual funds the Separation Theorem and the CAPM hold even with S-shape (or reverse S-shape) utility function, with change of wealth rather than total wealth and with CPT probability distortion. Thus, the CAPM is theoretically intact in the Rank Dependent Expected Utility (RDEU) framework. While the CAPM may not hold due to other factors (transaction costs, market segmentation, etc.), the experimental findings by themselves do not cause changes in the M-V efficiency analysis and the CAPM, as long as mutual funds exist in the market.

To sum up, though experimental findings open a new way of thinking on financial theory, most financial models, albeit not all of them, are robust even with these experimental findings. However, in some extreme cases, experimental evidence may explain phenomena, which cannot be explained by rational models. If investors follow some bounded rational behavior, booms and crashes in the stock market may be obtained (see Levy et al. 2000) even though there is no classical economic explanation for such a stock market behavior.

Notes

1. The analysis of the various alternate preferences of Fig. 2 is taken from Levy and Levy (2002).
2. For example, Markowitz argues that individuals with the Friedman and Savage utility function and wealth in the convex region would wish to take large symmetric bets, which is in contradiction to empirical observation.
3. For PSD see Levy and Wiener (1998) and Levy (1998) and Levy (1998). For MSD see Levy and Levy (2002a).
4. A quote from Mr. Lipson, a president of Horizon Financial Advisers, see *Wall Street Journal*, 10 April 1992 (an article by Ellen E. Schultz).

Acknowledgements The author acknowledges the financial support of the Krueger Center of Finance.

References

- Allais, M. 1953. Le Comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica* 21: 503–564.
- Arrow, K. 1965. *Aspects of the theory of risk bearing*. Helsinki: Yrjö Jahnssonin Säätiö.
- Arrow, K.J. 1971. *Essays in the theory of risk bearing*. Chicago: Markham Publishing Company.
- Benartzi, S., and R. Thaler. 1995. Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics* 110 (1): 73–92.
- . 1999. Risk aversion or myopia? Choices in repeated gambles and retirement investments. *Management Science* 45 (3): 364–381.
- . 2001. Naive diversification strategies in retirement savings plans. *American Economic Review* 91 (1): 79–98.
- Birnbaum, M.H. 1997. Violations of monotonicity in judgment and decision making. In *Choice, decision, and measurement: Essays in honor of R. Duncan Luce*, ed. A.A.J. Marley, 73–100. Mahwah: Erlbaum.
- Birnbaum, M.H., and W.R. McIntosh. 1996. Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes* 67: 91–110.
- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–654.

- Edwards, W. 1955. The prediction of decisions among bets. *Journal of Experimental Psychology* 50: 201–214.
- . 1962. Subjective probabilities inferred from decisions. *Psychology Review* 69: 109–135.
- Elton, E.J., and M.J. Gruber. 1984. *Modern portfolio theory and investment analysis*. 2nd ed. New York: Wiley.
- Fama, E.F. 1965. The behavior of stock market prices. *Journal of Business* 38: 34–105.
- . 1991. Efficient capital markets: II. *Journal of Finance* 46: 1575–1617.
- Fishburn, P.C. 1978. On Handa's "new theory of cardinal utility" and the maximization of expected return. *Journal of Political Economy* 86: 321–324.
- Friedman, M., and L.J. Savage. 1948. The utility analysis of choices involving risk. *Journal of Political Economy* 56: 279–304.
- Funk, S.G., A. Rapoport, and L.V. Jones. 1979. Investing capital on safe and risky alternatives: An experimental study. *Journal of Experimental Psychology: General* 108: 415–440.
- Gordon, M.J., G.E. Paradis, and C.H. Rorke. 1972. Experimental evidence on alternative portfolio decision rules. *American Economic Review* 52: 107–118.
- Hadar, J., and W. Russell. 1969. Rules for ordering uncertain prospects. *American Economic Review* 59: 25–34.
- Hanoch, G., and H. Levy. 1969. The efficiency analysis of choices involving risk. *Review of Economic Studies* 36: 335–346.
- Kahneman, D., and A. Tversky. 1979. Prospect theory of decisions under risk. *Econometrica* 47 (2): 263–291.
- Kroll, Y., and H. Levy. 1992. Further tests of separation theorem and the capital asset pricing model. *American Economic Review* 82: 664–670.
- Kroll, Y., H. Levy, and A. Rapoport. 1988a. Experimental tests of mean-variance model for portfolio selection. *Organizational Behavior and Human Decision Processes* 42: 388–410.
- . 1988b. Experimental tests of the separation theorem and the capital asset pricing model. *American Economic Review* 78: 500–519.
- Leshno, M., and H. Levy. 2002. Preferred by "All" and preferred by "Most" decision makers: Almost stochastic dominance. *Management Science* 48: 1074–1085.
- Levy, H. 1992. Stochastic dominance and expected utility: Survey and analysis. *Management Science* 38 (4): 555–593.
- . 1997. Risk and return: An experimental analysis. *International Economic Review* 38: 119–149.
- . 1998. *Stochastic dominance: Investment decision making under uncertainty*. Boston: Kluwer Academic Publishers.
- Levy, M., and H. Levy. 2001. Testing the risk-aversion: A stochastic dominance approach. *Economics Letters* 71: 233–240.
- . 2002a. Prospect theory: Much ado about nothing? *Management Science* 48: 1334–1349.
- . 2002b. Experimental test of prospect theory value function. *Organizational Behavior and Human Decision Processes* 89: 1058–1081.
- Levy, H., and M. Levy. 2002c. Arrow-Pratt risk aversion, risk premium and decision weights. *Journal of Risk and Uncertainty* 25: 265–290.
- . 2004. Prospect theory and mean-variance analysis. *Review of Financial Studies* 17: 1015–1041.
- Levy, H., and Z. Wiener. 1998. Stochastic dominance and prospect dominance with subjective weighting functions. *Journal of Risk and Uncertainty* 16: 147–163.
- Levy, M., H. Levy, and S. Solomon. 2000. *Microscopic simulation of financial markets: From investor behavior to market phenomena*. San Diego: Academic Press.
- Lintner, J. 1965. Security prices, risk, and maximal gains from diversification. *Journal of Finance* 20: 587–615.
- Machina, M.J. 1994. Review of generalized expected utility theory: The rank-dependent model. *Journal of Economic Literature* 32: 1237–1238.
- Markowitz, H.M. 1952a. Portfolio selection. *Journal of Finance* 7: 77–91.
- . 1952b. The utility of wealth. *Journal of Political Economy* 60: 151–156.
- . 1959. *Portfolio selection*. New York: Wiley.

- . 1987. *Mean variance analysis, portfolio choice and capital markets*. New York: Basil Blackwell.
- Modigliani, F., and M.H. Miller. 1958. The cost of capital corporation finance and the theory of investment. *American Economic Review* 48: 261–297.
- Odean, T. 1998. Are investors reluctant to realize their losses. *Journal of Finance* 53: 1775–1798.
- Plott, C.R. 1979. The application of laboratory experimental methods to public choice. In *Collective decision making*, ed. C.S. Russell. Washington, DC: Resources for the Future.
- Pratt, J.W. 1964. Risk aversion in the small and in the large. *Econometrica* 32: 122–136.
- Prelec, D. 1998. The probability weighting function. *Econometrica* 66: 497–527.
- Preston, M.G., and P. Baratta. 1948. An experimental study of the auction-value of uncertain outcomes. *American Journal of Psychology* 61: 183–193.
- Quiggin, J. 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3: 323–343.
- . 1993. *Generalized expected utility theory: The rank dependent model*. Boston: Kluwer Academic Publishers.
- Rapoport, A. 1984. Effects of wealth on portfolio under various investment conditions. *Acta Psychologica* 55: 31–51.
- Ross, S.A. 1976. The arbitrage theory of capital asset pricing. *Econometrica* 32: 122–137.
- Samuelson, P.A. 1994. The long term case for equities. *Journal of Portfolio Management* 21: 15–24.
- Sharpe, W.F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- Smith, V.L. 1976. Experimental economics: Induced value theory. *American Economic Review Proceedings* 66: 274–279.
- . 1982. Microeconomic systems as an experimental science. *American Economic Review* 72: 923–955.
- Swalm, R.O. 1966. Utility theory – Insights into risk taking. *Harvard Business Review* 44: 123–136.
- Thaler, H. 1993. *Advances in behavioral finance*. New York: Russell Sage Foundation.
- Thaler, R.H. 1999. The end of behavioral finance. *Financial Analysts Journal* 55 (6): 12–17.
- Tobin, J. 1958. Liquidity preferences as behavior toward risk. *Review of Economic Studies* 25: 65–86.
- Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211: 453–480.
- . 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5: 297–323.
- Viscusi, W.K. 1989. Prospective reference theory: Toward an explanation of the paradoxes. *Journal of Risk and Uncertainty* 2: 235–264.
- von Neuman, J., and O. Morgenstern. 1944. *Theory of games and economic behavior*. Princeton: Princeton University Press.
- Wakker, P.P. 2003. The data of Levy and Levy, “Prospect theory: Much ado about nothing?” Actually support prospect theory. *Management Science* 48: 1334–1349.
- Wakker, P.P., I. Erev, and E.U. Weber. 1994. Comonotonic independence: The critical test between classical and rank-dependent utility theories. *Journal of Risk and Uncertainty* 9: 195–230.
- Wilde, L. 1980. In the case of laboratory experiments in economics. In *The philosophy of economics*, ed. J. Pitt. Dordrecht: Reidel.
- Yaari, M. 1987. The dual theory of choice under risk. *Econometrica* 55 (1): 95–115.



Merger and Acquisition: Definitions, Motives, and Market Responses

28

Jenifer Piesse, Cheng-Few Lee, Lin Lin, and Hsien-Chang Kuo

Abstract

Along with globalization, merger and acquisition has become not only a method of external corporate growth, but also a strategic choice of the firm enabling further strengthening of core competence. The mega-mergers in the last decades have also brought about structural changes in some industries, and attracted international attention. A number of motivations for merger and acquisition are proposed in the literature, mostly drawn directly from finance theory but with some inconsistencies. Interestingly, distressed firms are found to be predators and the market reaction to these is not always predictable. Several financing options are associated with takeover activity and are generally specific to the acquiring firm. Given the interest in the academic and business literature, merger and acquisition will continue to be an interesting but challenging strategy in the search for expanding corporate influence and profitability.

Keywords

Acquisition · Efficiency · LBO · Market reaction · Merger · Synergy · Takeover · Takeover financing · Takeover regulations · Wealth effect

J. Piesse

University of London, London, UK

e-mail: jenifer.piesse@kcl.ac.uk

C.-F. Lee (✉)

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: lee@business.rutgers.edu

L. Lin · H.-C. Kuo

National Chi-Nan University, Nantou, Taiwan, Republic of China

28.1 Introduction

Merger and acquisition (M&A) plays an important role in external corporate expansion, acting as a strategy for corporate restructuring and control. It is a different activity from internal expansion decisions, such as those determined by investment appraisal techniques. M&A can facilitate fast growth for firms and is also a mechanism for capital market discipline, which improves management efficiency and maximises private profits and public welfare.

28.2 Definition of “Takeover,” “Merger,” and “Acquisition”

Takeover, merger, and acquisition are frequently used synonymously, although there is clearly a difference in the economic implications of takeover and a merger (Singh 1971: Conventions and Definitions). An interpretation of these differences defines takeover and acquisition as activities by which acquiring firms can control more than 50% of the equity of target firms, whereas in a merger at least two firms are combined with each other to form a “new” legal entity. In addition, it has been suggested that imprudent takeovers accounted for more than 75% of corporate failure in listed manufacturing firms in the United Kingdom over the periods 1948–1960 and 1954–1960 (Singh 1971). In contrast, conglomerates resulting from mergers increased industry concentration during the same periods. Because of the different economic outcomes, distinguishing between these may be useful.

Other writers too have required a more careful definition of terms. Hampton (1989) claimed that “a merger is a combination of two or more businesses in which only one of the corporations survives” (Hampton 1989, p. 394). Using simple algebra, Singh’s (1971) concept of merger can be symbolized by $A + B = C$, whereas Hampton’s (1989) can be represented by $A + B = A$ or B or C . What is important is the different degrees of negotiating power of the acquirer and acquirer in a merger. Negotiating power is usually linked to the size or wealth of the business. Where the power is balanced fairly equally between two parties, a new enterprise is likely to emerge as a consequence of the deal. On the other hand, in Hampton’s (1989) definition, one of the two parties is dominant.

The confusion worsens when the definition replaces the word ‘negotiating power’ with ‘chief beneficiary’ and ‘friendliness’ (Stallworthy and Kharbanda 1988). This claim is that the negotiating process of mergers and acquisitions is usually ‘friendly’ where all firms involved are expected to benefit, whereas takeovers are usually hostile and proceed in an aggressive and combative atmosphere. In this view, the term ‘acquisition’ is interchangeable with ‘merger’, while the term ‘takeover’ is closer to that of Singh’s (1971).

Stallworthy and Kharbanda (1988, pp. 26, 68) are not so concerned with the terminology and believed that it is meaningless to draw a distinction in practice. They also claim that the financial power of firms involved is the real issue. If one party is near bankruptcy, this firm will face very limited options and play the role of target in any acquisition activity. Rees (1990) disagrees and argues that is

unnecessary to distinguish between terms because they arise from a similar legal framework in the United Kingdom.

28.3 Motives for Takeover

The rationale for takeover activity has been discussed for many years (see Brealey et al. 2001, p. 641; Ross et al. 2002, p. 824). Unfortunately, no single hypothesis is sufficient to cover all takeovers and it is because the motives for takeovers are very complicated that it is useful to develop some framework to explain this activity. Of the numerous explanations available, the following are the most common in the literature, which has prompted the development of some hypotheses to explain takeover activities. Of these, eight broad reasons for takeover have emerged:

- Efficiency Theory
- Agency Theory
- Free Cash Flow Hypothesis
- Market Power Hypothesis
- Diversification Hypothesis
- Information Hypothesis
- Bankruptcy Avoidance Hypothesis
- Accounting and Tax Effects

Each are discussed in the next section, and clearly many are not mutually exclusive.

28.3.1 Efficiency Theories

Efficiency theories include differential efficiency theory and inefficiency management theory. Differential efficiency theory suggests that, providing firm *A* is more efficient than firm *B* and both are in the same industry, *A* can raise the efficiency of *B* to at least the level of *A* through takeover. Inefficiency management theory indicates that information about firm *B*'s inefficiency is public knowledge, and not only firm *A* but also the controlling group in any other industry can bring firm *B*'s efficiency to the acquirer's own level through takeover. These two theories are similar in viewing takeover as a device to improve the efficiency problem of the target firm. However, one difference is that firm *B* is not so inefficient that it is obvious to the firms in different industries in the first, but it is in the second. Thus, Copeland and Weston (1988) concluded that differential efficiency theory provides a theoretical basis for horizontal takeovers while inefficiency management theory supports conglomerate takeovers.

In the economics literature, efficiency assumes the optimal allocation of resources. A firm is Pareto efficient if there is no other available way to allocate resources without a detrimental effect elsewhere. However, at the organizational

level, a firm cannot be efficient unless all aspects of its operations are efficient. Therefore, in this literature a simplified but common definition of efficiency is that 'a contract, routine, process, organization, or system is efficient in this sense if there is no alternative that consistently yields unanimously preferred results' (Milgrom and Roberts 1992, p. 24). According to this definition, to declare a firm inefficient requires that another is performing better in similar circumstances, thus avoiding the problem of assessing the intangible parts of a firm as part of an efficiency evaluation.

The idea of efficiency in the takeover literature arises from the concept of synergy, which can be interpreted as a result of combining and coordinating the good parts of the companies involved as well as disposing of those that are redundant. Synergy occurs where the market value of the two merged firms is higher than the sum of their individual values. However, as Copeland and Weston (1988, p. 684) noted, early writers such as Myers (1968) and Schall (1972), were strongly influenced by Modigliani–Miller model (MM) (1958), who argued that the market value of two merged companies together should equal the sum of their individual values. This is because the value of a firm is calculated as the sum of the present value of all investment projects and these projects are assumed to be independent of other firms' projects. But this Value Additivity Principle is problematic when applied to the valuation of takeover effects. The main assumption is very similar to that required in the MM models, including the existence of a perfect capital market and no corporate taxes. These assumptions are very unrealistic and restrict the usefulness of the Value Additivity Principle in practice. In addition, the social gains or losses are usually ignored in those studies. Apart from those problems, the value creation argument has been supported by empirical studies. For example, Seth (1990) claimed that in both unrelated and related takeovers, value can be created to the same degree.

Synergy resulting from takeover can be achieved in several ways. It normally originates from the better allocation of resources of the combined firm, such as the replacement of the target's inefficient management with a more efficient one (Ross et al. 2002, p. 826) and the disposal of redundant and/or unprofitable divisions. Such restructuring usually has a positive effect on market value. Leigh and North (1978) found that this post-takeover and increased efficiency resulted from better management practices and more efficient utilisation of existing assets.

Synergy can also be a consequence of "operational" and "financial" economies of scale through takeovers (see Brealey et al. 2001, p. 641; Ross et al. 2002, p. 825). Operational economies of scale brings about the 'potential reductions in production or distribution costs' (Jensen and Ruback 1983, p. 611) and financial economies of scale includes lower marginal cost of debt and better debt capacity. Other sources of synergy are achieved through oligopoly power and better diversification of corporate risk. Many sources of synergy have been proposed and developed into separate theories to be discussed in later sections.

Finally, efficiency can be improved by the introduction of a new company culture through takeover. Culture may be defined as a set of secret and invisible codes that determines the behavior patterns of a particular group of people, including their way of thinking, feeling, and perceiving everyday events. Therefore, it is rational to

speculate that a successful takeover requires the integration of both company cultures in a positive and harmonious manner. Furthermore, the stimulation of new company culture could itself be a purpose of takeover, as Stallworthy and Kharbanda (1988) noted, and the merger of American Express and Shearson Loeb Rhoades (SLR) is a good example of this.

However, disappointing outcomes occur when a corporate culture is imposed on another firm following takeover conflict. This can take some time and the members of both organisations may take a while to adjust. Unfortunately, the changing business environment does not allow a firm much time to manage this adjustment and this clash of corporate cultures frequently results in corporate failure. Stallworthy and Kharbanda (1988, p. 93) found that, “it is estimated that about one-third of all acquisitions are sold off within 5 years. . . the most common cause of failure is a clash of corporate cultures, or ‘the way things are done round here’.”

28.3.2 Agency Theory

Agency theory is concerned with the separation of interests between company owners and managers (Jensen and Meckling 1976). The main assumption of agency theory is that principals and agents are all rational and wealth-seeking individuals who are trying to maximize their own utility functions. In the context of corporate governance, the principal is the shareholder and the agent is the directors/senior management. The neoclassical theory of the firm assumes profit maximization is the objective, but more recently in the economics literature other theories have been proposed, such as satisficing behavior on the part of managers, known as behavioral theories of the firm. Since management in a diversified firm does not own a large proportion of the company shares, they will be more interested in the pursuit of greater control, higher compensation, and better working conditions at the expense of the shareholders of the firm. The separation of ownership and control within a modern organization also makes it difficult and costly to monitor and evaluate the efficiency of management effectively. This is known as “moral hazard” and is pervasive both in market economies and other organizational forms. Therefore, managing agency relationships is important in ensuring that firms operate in the public interest.

A solution to the agency problem is the enforcement of contractual commitments with an incentive scheme to encourage management to act in shareholders’ interests. It can be noted that management compensation schemes vary between firms as they attempt to achieve different corporate goals. One of the most commonly used long-term remuneration plans is to allocate a fixed amount of company shares at a price fixed at the beginning of a multiyear period to managers on the basis of their performance at the end of the award period. By doing so, managers will try to maximize the value of the shares in order to benefit from this bonus scheme, thereby maximizing market value of the firm. Therefore, the takeover offer initiated by the firm with long-term performance plans will be interpreted by the market as good news since its managers’ wealth is tied to the value of the firm, a situation parallel to

that of shareholders. Empirically, it can be observed that “the bidding firms that compensate their executives with long-term performance plans, experience a significantly favorable stock market reaction around the announcements of acquisition proposals, while bidding firms without such plans experience the opposite reaction” (Tehrani et al. 1987, p. 74). Appropriate contracting can certainly reduce agency problems.

However, contracting may be a problem where there is information asymmetry. Managers with expertise can provide distorted information or manipulate reports to investors with respect to an evaluation of their end of period performance. This phenomenon is “adverse selection” and reflects information asymmetry in markets, a problem that is exacerbated when combined with moral hazard. Milgrom and Roberts (1992, p. 238) concluded that “the formal analysis of efficient contracting when there is both moral hazard and adverse selection is quite complex.”

Another solution may be takeover. Samuelson (1970, p. 505) claimed that “takeovers, like bankruptcy, represent one of Nature’s methods of eliminating deadwood in the struggle for survival.” An inefficient management may be replaced following takeover, and according to Agrawal and Walkling (1994), encounters great difficulty in finding an equivalent position in other firms without considerable gaps in employment. In this way, takeover is regarded as a discipline imposed by the capital markets. Jensen and Ruback (1983) claimed that the threat of takeover will effectively force managers to maximize the market value of the firm as shareholders wish, and thus eliminate agency problems, or their companies will be acquired and they will lose their jobs. This is consistent with the observations of some early writers such as Manne (1965).

Conversely, takeover could itself be the source of agency problems. Roll’s (1986) hubris hypothesis suggests that the management of the acquirer is sometimes over-optimistic in evaluating potential targets because of information asymmetry, and in most cases, because of their own misplaced confidence about their ability to make good decisions. Their over-optimism eventually leads them to pay higher bid premiums for potential synergies, unaware that the current share price may have fully reflected the real value of this target. In fact, acknowledging that takeover gains usually flow to shareholders, while employee bonuses are usually subject to the size of the firm, managers are encouraged to expand their companies at the expense of shareholders (Malatesta 1983). The hubris theory suggests that takeover is both a cause of and a remedy for agency problems. Through takeover, management not only increase their own wealth but also their power over richer resources, as well as an increased view of their own importance. But a weakness in this theory is the assumption that efficient markets do not notice this behavior. According to Mitchell and Lehn (1990), stock markets can discriminate between “bad” and “good” takeovers and bad bidders usually turn to be good targets later on. These empirical results imply that takeover is still a device for correcting managerial inefficiency, if markets are efficient. Of course, good bidders may be good targets too, regardless of market efficiency. When the market is efficient, a growth-oriented company can become an attractive target for more successful or bigger companies who wish to expand their business. When firms are inefficient, a

healthy bidder may be mistaken for a poor one and the resulting negative reaction will provide a chance for other predators to own this newly combined company. In these cases, the treatment directed towards target management may be different since the takeover occurs because of good performance not poor. In either case, Mitchell and Lehn (1990) admitted on the one hand that managers' pursuit of self-interest could be a motive for takeover but on the other they still argue that this situation will be corrected by the market mechanism.

28.3.3 Free Cash Flow Hypothesis

Closely connected to agency theory is the free cash flow hypothesis. Free cash flow is defined as "cash flow in excess of that required to fund all projects that have positive net present values when discounted at the relevant cost of capital (Jensen 1986, p. 323)." Free cash flow is generated from economic rents or quasi rents. Jensen (1986) argued that management is usually reluctant to distribute free cash flow to shareholders primarily because it will substantially reduce the company resources under their control while not increasing their own wealth since dividends are not their personal goal but bonus schemes. However, the expansion of the firm is a concern in management remuneration schemes so that free cash flow can be used to fund takeover, and thus grow the company. In addition, because fund-raising in the market for later investment opportunities puts management under the direct gaze of the stock market, there is an incentive for management to hold some free cash flow or internal funds for such projects (Rozeff 1982; Easterbrook 1984). Consequently, managers may prefer to retain free cash to grow the company by takeover, even though sometimes the returns on such projects are less than the cost of capital. This is consistent with the empirical results suggesting that organizational inefficiency and over-diversification in a firm are normally the result of managers' intention to expand the firm beyond its optimal scale (Gibbs 1993). Unfortunately, according to agency theory, managers' behaviors with respect to the management of free cash flow are difficult to monitor.

Compared with using free cash in takeovers, holding free cash flow too long may also not be optimal. Jensen (1986) found that companies with a large free cash flow become an attractive takeover target. This follows simply because takeover is costly and acquiring companies prefer a target with a good cash position to reduce the financial burden of any debt that is held now or with the combined company in the future. Management would rather use up free cash flow (retention) for takeovers than keep it within the firm. However, Gibbs (1993, p. 52) claims that free cash flow is only a "necessary condition for agency costs to arise, but not a sufficient condition to infer agency costs." In practice, some methods such as reinforcement of outside directors' power have also been suggested as a way to mitigate the potential agency problems when free cash flow exists within a firm. Apart from this legal aspect, management's discretion is also conditioned by fear of corporate failure. In a full economic analysis, an equilibrium condition must exist while the marginal bankruptcy costs equal the marginal benefits that management can gain through projects.

Again, the disciplinary power of the market becomes a useful weapon against agency problem regarding the management of free cash flow.

28.3.4 Market Power Hypothesis

Market power may be interpreted as the ability of a firm to control the quality, price, and supply of its products as a direct result of the scale of their operations. Because takeover promises rapid growth for the firm, it can be viewed as a strategy to extend control over a wider geographical area and enlarge the trading environment (Leigh and North 1978, p. 227). Therefore the market power hypothesis can serve as an explanation for horizontal and vertical takeovers.

Economic theory of oligopoly and monopoly identifies the potential benefits to achieving market power, such as higher profits and barriers to entry. The market power hypothesis therefore explains the mass of horizontal takeovers and the increasing industrial concentration that occurred during the 1960s. For example, in the United Kingdom, evidence shows that takeovers “were responsible for a substantial proportion of the increase in concentration over the decade 1958–1968 (Hart and Clarke 1980, p. 99).”

This wave of horizontal takeovers gradually decreased during recent years, primarily because of antitrust legislation introduced by many countries to protect the market from undue concentration and subsequent loss of competition that results. Utton (1982, p. 91) noted that tacit collusion can create a situation in which only a few companies with oligopolistic power can share the profits by noncompetitive pricing and distorted utilization and distribution of resources at the expense of society as a whole. In practice, antitrust cases occur quite frequently. For example, one of the most famous antitrust examples in the early 1980s was the merger of G. Heileman and Schlitz, the sixth and fourth largest companies in the US brewing industry. The combined company would have become the third largest brewer in the United States, but this was prohibited by the Department of Justice on anti-competitive grounds. Similarly, in the United Kingdom, GEC’s bid for Plessey was blocked by the Monopolies and Mergers Commission (MMC) in 1989 on the grounds of weakening price competition and Ladbroke’s acquisition of Coral in 1998 was stopped for the same reason. At an international level, the US and European antitrust authorities were ready to launch detailed investigations in 1998 into the planned takeover of Mobil, the US oil and gas group, by Exxon, the world’s largest energy group. More recently, irritated by antitrust lawsuits against him, Bill Gates of Microsoft accused the US government of attempting to destroy his company. However, horizontal takeovers are not the only target of the antitrust authorities and vertical and conglomerate takeovers are also of concern. This is because a “large firm’s power over prices in an individual market may no longer depend on its relative size in that market but on its overall size and financial strength (Utton 1982, p. 90).”

28.3.5 The Diversification Hypothesis

The diversification hypothesis provides a theoretical explanation for conglomerate takeovers. The diversification of business operations, i.e. the core businesses of different industries has been broadly accepted as a strategy to reduce risk and stabilise future income flows. It is also an approach to ensure survival in modern competitive business environments. In the United Kingdom, Goudie and Meeks (1982) observed that more than one-third of listed companies experiencing takeover in mainly manufacturing and distribution sectors during 1949–1973 could be classified as conglomerates. Since then, conglomerate takeover has become widespread as an approach to corporate external growth (Stallworthy and Kharbanda 1988; Weston and Brigham 1990).

Although different from Schall's (1971, 1972) Value Additivity Principle, Lewellen's (1971, 1972) coinsurance hypothesis provides a theoretical basis for corporate diversification. This argues that the value of a conglomerate will be greater than the sum of the value of the individual firms because of the decreased firm risk and increased debt capacity (see also Ross et al. 2002, pp. 828–829, 830–833). Appropriate diversification can effectively reduce the probability of corporate failure, which facilitates conglomerate fund raising and increases market value. Kim and McConnell (1977) noted that the bondholders of conglomerates were not influenced by the increased leverage simply because the default risk is reduced. This result remains valid even when takeovers were financed by increased debt. Takeover can also result in an increased debt capacity as the merged firm is allowed to carry more tax subsidies, and according to the MM Proposition (1958, 1963), the tax shield provided by borrowings is a dominant factor in firm valuation. In summary, the potentially higher tax deductions, plus the reduced bankruptcy costs, suggest that conglomerates will be associated with higher market values after takeovers.

Corporate diversification can also improve a firm's overall competitive ability. Utton (1982) stated that large diversified firms use their overall financial and operational competence to prevent the entry of rivals. One way to achieve this is through predatory pricing and cross subsidization, both of which can effectively form an entry barrier into the particular industry, and force smaller existing competitors out of the market. Entry via takeover reveals the inefficiency of incumbents as entry barriers are successfully negotiated. McCardle and Viswanathan (1994, p. 5) predicted that the stock prices of such companies should suffer. In fact, many writers had discussed this "build or buy" decision facing potential entrants (Fudenberg and Tirole 1986; Harrington 1986; Milgrom and Roberts 1982). McCardle and Viswanathan (1994) used game theory to model the market reaction to direct/indirect entry via takeover. From these game theoretic models, there are indications that corporate diversification will not cause an increase in market value for the newly combined firm as opposed to Lewellen's (1971, 1972) coinsurance hypothesis, weakening the justification of diversification as a motive for takeover.

28.3.6 The Information Hypothesis

The information hypothesis stresses the signaling function of many firm-specific financial policies and announcements. It argues that such announcements are trying to convey information still not publicly available to the market and predict a revaluation of the firm's market value, assuming efficient markets. Takeovers have the same effect. Both parties release some information in the course of takeover negotiations and the market may then revalue previously undervalued shares.

This hypothesis has been supported by numerous event studies, demonstrating substantial wealth changes of bidders and targets (see the summary paper of Jensen and Ruback 1983). Sullivan et al. (1994, p. 51) also found that the share prices of the firms involved in takeover "are revalued accordingly as private information is signaled by the offer medium that pertains to the target firm's stand-alone value or its unique synergy potential." Bradley et al. (1983) proposed two alternative forms of the information hypothesis. The first is referred to as the "kick-in-the-pants" hypothesis, which claims that the revaluation of share price occurs around the firm-specific announcements because management is expected to accept higher-valued takeover offers. The other is the "sitting-on-a-gold-mine" hypothesis asserting that bidder management is believed to have superior information about the current status of targets so that premiums would be paid. These two explanations both stress that takeover implies information sets which are publicly unavailable and favor takeover proposals. It is also noted that these two forms of information hypothesis are not mutually exclusive, although not all empirical research supports the information hypothesis (Bradley 1980; Bradley et al. 1983; Dodd and Ruback 1977; Firth 1980; Van Horne 1986).

Finally, the information hypothesis is only valid where there is strong-form market efficiency. Ross's signaling hypothesis (1977) points out that management will not give a false signal if its marginal gain from a false signal is less than its marginal loss. Therefore, it cannot rule out the possibility that management may take advantage of investors' naivety to manipulate the share price. The information hypothesis only suggests that takeover can act as a means of sending unambiguous signals to the public about the current and future performance of the firm, but does not take management ethics into account.

28.3.7 The Bankruptcy Avoidance Hypothesis

The early economic literature did not address bankruptcy avoidance as a possible motivation for takeover, largely because of the infrequent examples of the phenomenon. However, some writers (for example, Altman 1971) suggest the potential link between takeover and bankruptcy in financial decisions. Stiglitz (1972) argued that enterprises can avoid the threat of either bankruptcy or takeover through appropriately designed capital structures and regards takeover as a substitute for bankruptcy. Shrieves and Stevens (1979) also examined this relationship between takeover and

bankruptcy as a market disciplining mechanism and found that a carefully timed takeover can be an alternative to bankruptcy.

However, intuition suggests that financially unhealthy firms are not an attractive target to potential predators. One way to resolve this dilemma is to consider the question from the bidder and target perspectives separately. To acquirers, the immediate advantages of a distressed target are the discounted price and lack of competition from other predators in the market. Much management time and effort is involved in searching and assessing targets, as well as the negotiation and funding process. This is much less for a distressed target than for a healthy one (Walker 1992, p. 2). In addition, there may be tax benefits as well as the expected synergies. From the target shareholders' viewpoint, the motivation is more straightforward. Pastena and Ruland (1986, p. 291) noted that "with respect to the merger/bankruptcy choice, shareholders should prefer merger to bankruptcy because in a merger the equity shareholders receive stock while in bankruptcy they frequently end up with nothing."¹ However, while the bankruptcy avoidance hypothesis can be justified from the bidder and target shareholder perspectives, it fails to take the agency problem into account. Ang and Chua (1981) found that managers of a distressed company tended to stay in control if there was a rescue package or the firm was acquired.

However, not all distressed firms welcome acquisition as a survival mechanism and Gilson (1989) suggested that agency problems may not be the reason for the management of a distressed firm to reject a takeover offer. Managers dismissed from failing firms that filed for bankruptcy or private debt restructuring during 1979–1984, were still unemployed 3 years later, while those still in post were on reduced salary and a scaled-down bonus scheme (Gilson and Vetsuypens 1993). Clearly, bankruptcy is costly to managers as well as other stakeholders.

If takeover can serve as a timely rescue for distressed companies, bankrupt firms present similar characteristics as distressed targets. In a two-country study, Peel and Wilson (1989, p. 217) found that in the United Kingdom, factors associated with corporate failure are similar to those in acquired distressed firms. These include longer time lags in reporting annual accounts, a going-concern qualification, and a high ratio of directors' to employees' remuneration, while neither company size or ownership concentration was important. However, in the United States, different factors were identified, with the differences attributed to the variation between the UK and US business environment.

Finally, although the benefits of acquiring distressed companies have been identified, Walker (1992) argued that there are economic advantages to acquiring distressed firms after their insolvency, as many problems will be solved by receivers at the time they are available for sale. Clearly, this weakens the validity of the bankruptcy avoidance hypothesis.

28.3.8 Accounting and Tax Effects

Profiting from accounting and tax treatments for takeover could be another factor influencing the takeover decision. Two accounting methods are at issue: the pooling

of interests and the purchase arrangements. Copeland and Weston (1988) defined them as follows,

In a pooling arrangement the income statements and balance sheets of the merging firms are simply added together. On the other hand, when one company purchases another, the assets of the acquired company are added to the acquiring company's balance sheet along with an item called goodwill. . . [which is] the difference between the purchase price and the book value of the acquired company's assets. . . [and, by regulation, should] be written off as a charge against earnings after taxes in a period not to exceed 40 years. (Copeland and Weston 1988, p. 365)

Thus, the difference between the pooling and purchase methods lies in the treatment of goodwill, which is not recognized in the former but is in the latter. Not surprisingly, these two accounting treatments have different effects on company's postmerger performance. It is observed that "when the differential is positive (negative), the pooling (purchase) method results in greater reported earnings and lower net assets for the combined entity. . . the probability of pooling (purchase) increases with increases (decreases) in the differential (Robinson and Shane 1990, p. 26)." After much debate, the pooling method was prohibited in the United States in 2001, which abolishes the accounting effects as a reason for merger and acquisition.

However, takeover can be motivated by tax considerations on the part of the owner. For example, a company paying tax at the highest rate may acquire an unsuccessful company in an attempt to lower its overall tax payment (Ross et al. 2002, p. 827). This may extend to country effects in that a firm registered in a low-corporate tax region will have a reduced tax liability from assets transferred associated with a takeover. The globalization of business increases the opportunity for cross-border takeovers, which not only reflect the tax considerations but have longer-term strategic implications.

28.4 Methods of Takeover Financing and Payment

A takeover can be financed through borrowings (cash) or the issue of new equity, or both (see Brealey et al. 2001, pp. 645–648; Ross et al. 2002, pp. 835–838). The sources of debt financing include working capital, term debt, vendor take-back, subordinated debt, and government contributions, while equity financing consists of mainly preferred and common shares, and also retained earnings (Albo and Henderson 1987). The financing decision is specific to the acquiring firm and considerations such as equity dilution, risk policy, and current capital structure. Of course, the interrelation between the participants in the capital markets and the accessibility of different sources of financing is critical to any financing decision.

In debt financing, borrowers' credibility is the main concern of the providers of capital in determining the size and maturity of the debt. Some additional investigation may be conducted before a particular loan is approved. For example, lenders will be interested in the value of the underlying tangible assets to which an asset-

based loan is tied to the capacity and steadiness of the cash flow stream of the borrower for a cash flow loan.

Equity financing can be divided into external and internal elements. External equity financing through the stock market is bad news as issuing new equity implies an overvalued share price, according to the signaling hypothesis. In contrast, debt financing is regarded as good news because increasing the debt-to-equity ratio of a firm implies managers' optimism about future cash flows and reduced agency problems. Therefore, debt financing is welcomed by the stock market as long as it does not raise gearing levels too much.

Reserves are an internal source of equity financing, and is the net income not distributed to shareholders or used for investment projects, which then become part of owners' future accumulated capital. Donaldson (1961) and Myers (1984) suggest that a firm prefers reserves over debt and external equity financing because it is not subject to market discipline. This ranking of preferences is called the "the pecking order theory." However, given possible tax advantages, debt financing increases the market value of the firm to the extent that the marginal gain from borrowings is equal to the marginal expected loss from bankruptcy. The contradictory implications arising from these hypotheses results from the fundamentally different assumptions on which they are based. The pecking order theory of funding preference emphasizes agency theory, while the static trade-off argument that determines optimal capital structure assumes that managers' objectives are to maximize the market value of the firm. As to external equity financing, since this is a negative signal to the market and subject to unavoidable scrutiny, it is the last choice of funding for predators.

However, distressed acquirers have fewer options. Firstly, they may not have sufficient reserves for a takeover and may have to increase their already high gearing levels. They are also unwilling to issue new stocks, as this will jeopardize the current share price. Alternatively, they can initiate takeovers after resolving some problems through a voluntary debt restructuring strategy. Studies on the relationship between troubled firms and their debt claimants suggest that distressed firms have a better chance of avoiding corporate failure if the restructuring plan fits their current debt structure (Asquith et al. 1994; Brown et al. 1993; Gilson et al. 1990; John et al. 1992). Finally, distressed acquirers can finance takeovers by selling off part of the firm's assets. Brown et al. (1994) noted that such companies can improve the efficiency of their operations and management and repay their debts by partial sale of assets.

A growing literature on method of takeover payment shows the existence of a relationship between methods of takeover payment and of financing for takeover. Most of the research focuses on the common stock exchange offer and cash offer (Sullivan et al. 1994; Travlos 1987). Those studies imply that wealthy firms initiate a cash offer but distressed ones prefer an all-share bid. However, it is not only the users that differentiate cash offers from all-share offers. As Fishman (1989, p. 41) pointed out, "a key difference between a cash offer and a (risky) securities offer is that a security's value depends on the profitability of the acquisition, while the value of cash does not." Therefore, it is reasonable to assume that the "costs" of using a cash offer are lower than those using an all-share offer, given conditions of information

asymmetry. In addition, cash offers are generally accepted in “preempt competition,” in which high premiums must be included in cash offers to “ensure that sufficient shares are tendered to obtain control (Hirshleifer and Titman 1990, p. 295).”

28.5 Market Reaction to Acquiring Firms

Compared to research on the wealth effects of takeover on target shareholders, research on the effects on bidder shareholders is limited. Moreover, the results for target shareholders are more consistent (see Brealey et al. 2001, pp. 652, 657; Ross et al. 2002, pp. 842–845) whereas those for bidder shareholders are still inconclusive. Halpern (1983, pp. 306–308) noted

The one consistent finding for all merger and takeover residual studies is the presence of large and significant positive abnormal returns and CAR's for the target firm's shareholders regardless of the definition of the event date. . . From the discussion of the abnormal returns to bidders it appears that tender offers are wealth maximising events. For mergers, the results are more ambiguous but leaning toward to the same conclusion.

Jensen and Ruback (1983), Langetieg (1978), Bradley (1980), Dodd (1980), and Malatesta (1983) use using event study methods to examine the market reaction to acquiring firms and concur with this result. More recently, Lin and Piesse (2004) argue that such ambiguities result from ignorance of the distortion effects of distressed acquirers in many samples and find the stock market reacts differently to nondistressed and distressed bidders, given semi-strong efficiency. Therefore, a sample that does not separate the two groups properly will inevitably result in confusing results, despite the noise that frequently accompanies takeover activity.

The long-term performance of acquiring firms is also a concern. Agrawal et al. (1992) found that after a failed bid, shareholders in the United States generally suffered a significant loss of about 10% over the following 5 years. Gregory (1997) came to the same conclusion despite known differences in the US and UK business environments, claiming this supported Roll's (1986) hubris hypothesis and agency theory.

28.6 Conclusion

Corporate mergers and acquisitions in industrialized economies are frequent and it is accepted that large mergers in particular have huge wealth redistribution effects as well as raising concerns for corporate governance and takeover codes. This activity is an useful corporate strategy, used by organizations to achieve various goals, and also acts as a mechanism for market discipline. A number of motivations for takeover have been discussed, although these are not mutually exclusive, while others are omitted altogether.

This chapter has reviewed studies on merger motives, financing and payment methods, wealth creation, and distribution between bidders' and target shareholders and the impact of takeovers on the competitors of predator and target companies (Chatterjee 1986; Song and Walking 2000). The growing scope for studies on takeover activity suggests that acquisition is an increasingly importance corporate strategy for changing business environments, and has implications for future industrial reorganization and the formation of new competitive opportunities.

Note

Especially in a competitive bidding situation, target shareholders usually receive a premium on the market price of their shares, although competition for distressed companies is rare.

Acknowledgements We would like to thank many friends in University of London (U.K.) and National Chi Nan University (Taiwan) for valuable comments. We also want to thank our research assistant Chiumei Huang for preparing the manuscript and proofreading several drafts of the manuscript. Last, but not least, special thanks go to the Executive Editorial Board of the *Encyclopedia in Finance* in Springer, who expertly managed the development process and superbly turned our final manuscript into a finished product.

References

- Agrawal, A., and R.A. Walkling. 1994. Executive careers and compensation surrounding takeover bids. *Journal of Finance* 49 (3): 985–1014.
- Agrawal, A., J. Jaffe, and G.N. Mandelker. 1992. The post-merger performance of acquiring firms: A re-examination of an anomaly. *Journal of Finance* 47 (4): 1605–1621.
- Albo, W.P., and R.A. Henderson. 1987. *Mergers and acquisitions of privately-held businesses*. Canadian Cataloguing in Publication Data.
- Altman, E.I. 1971. *Corporate bankruptcy in America*. New York: Lexington Books.
- Ang, J.S., and J.H. Chua. 1981. Corporate bankruptcy and job losses among top level managers. *Financial Management* 10 (5): 70–74.
- Asquith, P., R. Gertner, and D. Scharfstein. 1994. Anatomy of financial distress: An examination of junk bond issuers. *Quarterly Journal of Economics* 109 (3): 625–658.
- Bradley, M. 1980. Interfirm tender offers and the market for corporate control. *Journal of Business* 53 (4): 345–376.
- Bradley, M., A. Desai, and E.H. Kim. 1983. The rationale behind interfirm tender offers: Information or synergy. *Journal of Financial Economics* 11 (4): 183–206.
- Brealey, R.A., S.C. Myers, and A.J. Marcus. 2001. *Fundamentals of corporate finance*. 3rd ed. New York: McGraw-Hill.
- Brown, D.T., C.M. James, and R.M. Mooradian. 1993. The information content of distressed restructurings involving public and private debt claims. *Journal of Financial Economics* 33 (1): 93–118.
- . 1994. Asset sales by financially distressed firms. *Journal of Corporate Finance* 1 (2): 233–257.
- Chatterjee, S. 1986. Types of synergy and economic value: The impact of acquisitions on merging and rival firms. *Strategic Management Journal* 7 (2): 119–139.
- Copeland, T.E., and J.F. Weston. 1988. *Financial theory and corporate policy*. 3rd ed. Reading: Addison-Wesley.

- Dodd, P. 1980. Merger proposals, management discretion and stockholder wealth. *Journal of Financial Economics* 8: 105–137.
- Dodd, P., and R. Ruback. 1977. Tender offers and stockholder returns: An empirical analysis. *Journal of Financial Economics* 5 (3): 351–374.
- Donaldson, G. 1961. *Corporate debt capacity: A study of corporate debt policy and the determination of corporate debt capacity*. Boston: Division of Research, Harvard Graduate School of Business Administration.
- Easterbrook, F.H. 1984. Two agency-cost explanations of dividends. *American Economic Review* 74 (4): 650–659.
- Firth, M. 1980. Takeovers, shareholder returns, and the theory of the firm. *Quarterly Journal of Economics* 94 (2): 235–260.
- Fishman, M.J. 1989. Preemptive bidding and the role of the medium of exchange in acquisitions. *Journal of Finance* 44 (1): 41–57.
- Fudenberg, D., and J. Tirole. 1986. A “signal-jamming” theory of predation. *RAND Journal of Economics* 17 (3): 366–376.
- Gibbs, P.A. 1993. Determinants of corporate restructuring: The relative importance of corporate governance, takeover threat, and free cash flow. *Strategic Management Journal* 14 (special issue): 51–68.
- Gilson, S.C. 1989. Management turnover and financial distress. *Journal of Financial Economics* 25 (2): 241–262.
- Gilson, S.C., and M.R. Vetsuypens. 1993. CEO compensation in financially distressed firms: An empirical analysis. *Journal of Finance* 48 (2): 425–458.
- Gilson, S.C., K. John, and L.H.P. Lang. 1990. Troubled debt restructurings: An empirical study of private reorganisation of firms in default. *Journal of Financial Economics* 27 (2): 315–353.
- Goudie, A.W., and G. Meeks. 1982. Diversification by merger. *Economica* 49 (196): 447–459.
- Gregory, A. 1997. An examination of the long run performance of UK acquiring firms. *Journal of Business Finance* 24 (7–8): 971–1002.
- Halpern, P. 1983. Corporate acquisitions: A theory of special cases? A review of event studies applied to acquisitions. *Journal of Finance* 38 (2): 297–317.
- Hampton, J.J. 1989. *Financial decision making: Concepts, problems, and cases*. 4th ed. Hoboken: Prentice-Hall.
- Harrington, J.E. 1986. Limit pricing when the potential entrant is uncertain of its cost function. *Econometrica* 54 (2): 429–438.
- Hart, P.E., and R. Clarke. 1980. *Concentration in British Industry 1935–75*. London: Cambridge University Press.
- Hirshleifer, D., and S. Titman. 1990. Share tendering strategies and the success of hostile takeover bids. *Journal of Political Economy* 98 (2): 295–324.
- Jensen, M.C. 1986. Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review* 76 (2): 323–329.
- Jensen, M.C., and W.H. Meckling. 1976. Theory of the firm: Managerial behaviour, agency cost and ownership structure. *Journal of Financial Economics* 3: 305–360.
- Jensen, M.C., and R.S. Ruback. 1983. The market for corporate control: The scientific evidence. *Journal of Financial Economics* 11 (1–4): 593–638.
- John, K., L.H.P. Lang, and J. Netter. 1992. The voluntary restructuring of large firms in response to performance decline. *Journal of Finance* 47 (3): 891–917.
- Kim, H., and J. McConnell. 1977. Corporate mergers and co-insurance of corporate debt. *Journal of Finance* 32 (2): 349–365.
- Langestieg, T.C. 1978. An application of a three-factor performance index to measure stockholder gains from merger. *Journal of Financial Economics* 6: 365–383.
- Leigh, R., and D.J. North. 1978. Regional aspects of acquisition activity in British manufacturing industry. *Regional Studies* 12 (2): 227–245.
- Lewellen, W.G. 1971. A pure financial rationale for the conglomerate merger. *Journal of Finance* 26 (2): 521–537.

- . 1972. Finance subsidiaries and corporate borrowing capacity. *Financial Management* 1 (1): 21–31.
- Lin, L., and J. Piesse. 2004. Financial risk assessment in takeover and the change of bidder shareholders' wealth. *International Journal of Risk Assessment and Management* 4 (4): 332–347.
- Malatesta, P.H. 1983. The wealth effect of merger activity and the objective functions of merging firms. *Journal of Financial Economics* 11: 155–181.
- Manne, H.G. 1965. Mergers and the market for corporate control. *Journal of Political Economy* 73 (2): 110–120.
- McCardle, K.F., and S. Viswanathan. 1994. The direct entry versus takeover decision and stock price performance around takeovers. *Journal of Business* 67 (1): 1–43.
- Milgrom, P., and J. Roberts. 1982. Limit pricing and entry under incomplete information: An equilibrium analysis. *Econometrica* 50 (2): 443–460.
- . 1992. *Economics, organisation and management*. Hoboken: Prentice-Hall.
- Mitchell, M.L., and K. Lehn. 1990. Do bad bidders become good targets? *Journal of Political Economy* 98 (2): 372–398.
- Modigliani, F., and M.H. Miller. 1958. The cost of capital, corporation finance, and the theory of investment. *American Economic Review* 48 (3): 261–297.
- . 1963. Corporate income taxes and the cost of capital. *American Economic Review* 53 (3): 433–443.
- Myers, S.C. 1968. Procedures for capital budgeting under uncertainty. *Industrial Management Review* 9 (3): 1–19.
- . 1984. The capital structure puzzle. *Journal of Finance* 39 (3): 575–592.
- Pastena, V., and W. Ruland. 1986. The merger bankruptcy alternative. *Accounting Review* 61 (2): 288–301.
- Peel, M.J., and N. Wilson. 1989. The liquidation/merger alternative: Some results for the UK corporate sector. *Managerial and Decision Economics* 10 (3): 209–220.
- Rees, B. 1990. *Financial analysis*. London: Prentice-Hall.
- Robinson, R.J., and P.B. Shane. 1990. Acquisition accounting method and bid premia for target firms. *Accounting Review* 65 (1): 25–48.
- Roll, R. 1986. The hubris hypothesis of corporate takeovers. *Journal of Business* 59 (2): 197–216.
- Ross, S.A. 1977. The determination of financial structure: The incentive signalling approach. *Bell Journal of Economics* 8 (1): 23–70.
- Ross, S.A., R.W. Westerfield, and J.F. Jaffe. 2002. *Corporate finance*. 6th ed. New York: McGraw-Hill.
- Rozeff, M. 1982. Growth, beta and agency costs as determinants of dividend payout ratios. *Journal of Financial Research* 5: 249–259.
- Samuelson, P. 1970. *Economics*. 8th ed. New York: McGraw-Hill.
- Schall, L.D. 1971. Firm financial structure and investment. *Journal of Financial and Quantitative Analysis* 6 (3): 925–942.
- . 1972. Asset valuation, firm investment, and firm diversification. *Journal of Business* 45 (1): 11–28.
- Seth, A. 1990. Value creation in acquisitions: A re-examination of performance issues. *Strategic Management Journal* 11: 99–115.
- Shrieves, R., and D. Stevens. 1979. Bankruptcy avoidance as a motive for merger. *Journal of Financial and Quantitative Analysis* 14 (3): 501–515.
- Singh, A. 1971. *Take-overs: Their relevance to the stock market and the theory of the firm*. Cambridge, UK: Cambridge University Press.
- Song, M.H., and R.A. Walking. 2000. Abnormal returns to rivals of acquisition targets: A test of "Acquisition Probability Hypothesis". *Journal of Financial Economics* 55: 143–171.
- Stallworthy, E.A., and O.P. Kharbanda. 1988. *Takeovers, acquisitions and mergers: Strategies for rescuing companies in distress*. London: Kogan Page.
- Stiglitz, J. 1972. Some aspects of the pure theory of corporate finance: Bankruptcies and takeover. *Bell Journal of Economics and Management Science* 3 (2): 458–482.

- Sullivan, M.J., M.R.H. Jensen, and C.D. Hudson. 1994. The role of medium of exchange in merger offers: Examination of terminated merger proposal. *Financial Management* 23 (3): 51–62.
- Tehrani, H., N. Travlos, and J. Waagele. 1987. The effect of long-term performance plans on corporate sell-off induced abnormal returns. *Journal of Finance* 42 (4): 933–942.
- Travlos, N.G. 1987. Corporate takeover bids, methods of payment, and bidding firms' stock returns. *Journal of Finance* 42 (4): 943–963.
- Utton, M.A. 1982. *The political economy of big business*. Oxford, UK: Martin Robertson.
- Van Horne, J.C. 1986. *Financial management and policy*. 7th ed. Hoboken: Prentice-Hall.
- Walker, I.E. 1992. *Buying a company in trouble: A practical guide*. Hants: Gower.
- Weston, J.F., and E.F. Brigham. 1990. *Essentials of managerial finance*. 9th ed. London: Dryden Press.



Multistage Compound Real Options: Theory and Application 29

William T. Lin, Cheng-Few Lee, and Chang-Wen Duan

Abstract

We explore primarily the problems encountered in multivariate normal integration and the difficulty in root-finding in the presence of unknown critical value when applying compound real call option to evaluating multistage, sequential high-tech investment decisions. We compared computing speeds and errors of three numerical integration methods. These methods, combined with appropriate root-finding method, were run by computer programs Fortran and Matlab. It is found that secant method for finding critical values combined with Lattice method and run by Fortran gave the fastest computing speed, taking only 1 s to perform the computation. Monte Carlo method had the slowest execution speed. It is also found that the value of real option is in reverse relation with interest rate and not necessarily positively correlated with volatility, a result different from that anticipated under the financial option theory. This is mainly because the underlying of real option is a nontraded asset, which brings dividend-like yield into the formula of compound real options.

In empirical study, we evaluate the initial public offering (IPO) price of a new DRAM chipmaker in Taiwan. The worldwide average sales price is the underlying variable and the average production cost of the new DRAM foundry is the exercise price. The twin security is defined to be a portfolio of DRAM manufacturing and packaging firms publicly listed in Taiwan stock markets. We estimate the dividend-like yield with two methods, and find the yield to be negative. The negative dividend-like yield results from the negative correlation between the newly constructed DRAM foundry and its twin security, implying

W. T. Lin · C.-W. Duan
Tamkang University, New Taipei, Taiwan, Republic of China

C.-F. Lee (✉)
Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA
e-mail: lee@business.rutgers.edu

the diversification advantage of a new generation of DRAM foundry with a relative low cost of investment opportunity. It has been found that there is only a 4.6% difference between the market IPO price and the estimated one.

Keywords

Average sales price · CAPM · Closed-form solution · Critical value · Dividend-like yield · DRAM chipmaker · DRAM foundry · Fortran · Gauss quadrature method · Investment project · IPO · Lattice method · Management flexibility · Matlab · Monte Carlo method · Multivariate normal integral · Non-traded asset · Real call option · Secant method · Strategic flexibility · Twin security · Uncertainty · Vector autoregression

29.1 Introduction

Since Brennan and Schwartz (1985) applied the options theory to the evaluation of natural resources investment projects, further researches in this area have focused on valuing specific forms of managerial or project flexibility and on determining how to optimally capture the full value of such flexibility, ignoring the net present value (NPV) framework. The biggest difference between real option and financial option models is that the underlying of real option is a nontraded asset, which is not reproducible. Thus, we cannot compute the value of real option under a risk-neutral framework.

McDonald and Siegel (1984, 1985) noted if the object of investment is a nontraded asset, its expected return will be lower than the equilibrium total expected rate of return required in the market from an equivalent-risk traded security. Thus there exists a rate of return shortfall. (d) in the real option pricing model, which is the difference between the security's expected rate of return (α_s) and the real growth rate of the underlying asset (α_v), rendering the pricing of nontraded assets and traded assets somewhat different. Trigeorgis (1993a, b) also observed that regardless of whether the underlying asset under valuation is traded or not, it may be priced in the world of systematic risk by substituting the real growth rate with certainty-equivalent rate.¹

In the rapidly developing economic environment, information acquired by the management of a business or investor is oftentimes incomplete. Management often needs to make investment decision under high uncertainty. In real world, a business frequently adjusts its investment decision in response to the uncertainty in the market. Traditional evaluation models for investment do not offer full management flexibility, which however may be remedied by the approach of real options. Keeley et al. (1996) indicate that a proper evaluation model must reflect the "high risk" and "multistage" nature of an investment project and capture the prospective profit growth of the firm. Trigeorgis (1994), Amram and Kulatilaka (1999), Copeland and Antikarov (2001), and McDonald (2002) suggested the use of real options approach for evaluation of

¹"Certainty-equivalent rate" means deducting risk premium from the original growth rate of underlying, which is the same as pricing in a risk neutral world.

investment decision. Relative to the net present value (NPV) approach, which employs the “one-dimensional” thinking of NPV being greater than zero or not, real options is a “two-dimensional” approach that concurrently captures the NPV of the hi-tech investment opportunity and the volatility contained in the uncertainty.

According to the economic growth theory of Schumpeter (1939), a normal and healthy economic system does not grow steadily along some fixed path, and creative destruction is the main reason for the disintegration of a fixed normal economic system. Schumpeter further observed that such creative destruction is brought about by technological innovations. Therefore, in a new industry, technological innovations, which induce more inventions, are the main cause of economic cycle. Innovations tend to attract investment activities that give the technology market effect and bring new profit opportunities. In industrialized countries, many studies have demonstrated that technological innovations drive long-run economic growth, improve productivity, and introduce new products to the market. It is no doubt that innovations bring growth and profit opportunities for businesses. In the U.S., the earnings of a high-tech firm often do not have a direct bearing on its stock price. More often than not, earnings and stock price of a firm move in opposite directions, indicating the value of a high-tech firm lies in innovations, and not in physical assets such as equipment and plant. Hence the valuation of the investment project of a continuously innovating high-tech firm with high profit is an interesting study. This chapter intends to explore whether innovations do bring big profit opportunities that coincide with the theory of Schumpeter (1939).

In the case study of ProMos, the company’s main product is DRAM. The DRAM products have strict requirements for process equipment and technology. For the DRAM industry, technological innovations are often illustrated in the process technology and equipment. The new DRAM foundry project of ProMos in 1996 fits the approach of real options. The costs of plant construction, operation, and R&D are very high. It is a capital-intensive investment project and the plant building will take several years. Thus, it involves a sequential multistage capital budgeting process (Trigeorgis 1994), characterized by the cash flow in initial stage of the project being small and that in later stages big. This project targets primarily in-the-money opportunity presented when the market undergoes rapid growth. A single-stage model is inadequate for this kind of project, while multistage real option is more appropriate for depicting the value of decision points throughout the project.

The model should be able to reflect the multistage and high-risk nature of high-tech investment. In simulation, we assume that the investment decision in each stage is made at the beginning of the stage, that when the value of real options is higher than the planned investment amount for the stage, the project will be implemented and continue until the end of the stage when the decision for the next stage is made. Thus, the real option for each stage is European style. As discussed above, the model used to evaluate an investment project involving high-tech industry must also remedy the fact that the investment project is a nontraded asset. Trigeorgis (1993a, b) handled the property of a nontraded asset with dividend-like yield, which is defined as the rate of return short fall. In the dividend-like yield, there exists a positive correlation between the underlying asset and twin security; if the dividend-like yield is positive, it suggests the positive correlation between the underlying asset

and twin security, implying poor diversification and high-opportunity costs of the new investment project; conversely, it implies diversification advantage and low-opportunity costs of the new investment project.

We extend the closed-form solution of Geske (1979) for two-stage compound financial options to a closed-form solution for multistage compound real options to depict the multistage, sequential nature of a high-tech investment project. We also examine the difference in valuation algorithm brought about by the inclusion of nontraded assets into the options theory. We also tackle the difficulty encountered in closed-form solution of multivariate normal integration and the nonlinear root-finding of critical value, and compare the computing speed and the degree of error reduction of different multivariate normal integration numerical procedures in combination with various critical value root-finding methods.

Finally, we study the new DRAM foundry investment case of ProMos to discuss how to select the underlying variable and twin security in an investment project that is a nontraded asset under the framework of real options. We will also estimate the exercise price and dividend-like yield, based on which, to determine of value of ProMos at the time of initial public offering (IPO) and carry out sensitivity analysis.

29.2 Real Options

The concept of real options was first proposed by Myers (1977), who observed that the assets of many firms, in particular investment projects with growth opportunity may be expressed as a call option. Real options apply the analytical framework of financial options, which take into account management flexibility and strategic flexibility overlooked in the traditional NPV approach, and consider the irreversibility and deferability of investment decision. Trigeorgis and Mason (1987) pointed out that corporate management frequently adopts decision mechanism with considerable flexibility when dealing with highly uncertain, large investment project. Thus the valuation of such project should include the traditional NPV₃ plus the options value derived from management flexibility, which is termed “*expanded NPV*”:

$$\text{Expanded NPV} = \text{Static NPV} + \text{Value of Real Option} \quad (1)$$

Thus, the higher the uncertainty and the longer the investment period, the greater the discount rate and the smaller the NPV, but the drop in NPV will be offset by the *value of real option* derived from management flexibility. That is why discount cash flow (DCF) based evaluation methods² are often questioned by researchers. The NPV approach is suitable for the valuation of fixed cash flow investments, such as bonds, but does not express well when the investment project has uncertain factors, such as

²Such as NPV and IRR.

strategic moves and subsequent investment opportunities. The real options approach can capture the value of latent profit opportunities brought about by such uncertainties.

In high-tech industries like biotechnology and semiconductor, the risks are high and cash flow is small, which may be even negative in the initial development of a new generation of technology. But when the product is accepted by the market and enters the mass production stage, cash flow is high and the stock price of the company often rises sharply. It is as if these businesses are out-of-the-money in their initial stage and become in-the-money in the mass production stage, bringing substantial prospective profit opportunity for investors who put money into the firm at the initial stage. Therefore the evaluation model for the high-tech investment project is different from the NPV method, which follows the theory of the higher the risk, the greater the discount rate and the smaller the NPV.

Luehrman (1998a, b) indicated that information required for evaluation of investment project using real options method is just expanded information for the traditional model and not difficult to obtain. Luehrman (1998a, b) also suggested that NPV method is a “one-dimensional” thinking that evaluates whether the NPV of the underlying asset is greater than zero, while real options method is a “two-dimensional” thinking, which takes into account the NPV of the underlying and the opportunity presented by uncertainties. Thus in the evaluation of investment project with high uncertainty, the latter offers a better decision-making approach than other methods. Through the concept of flexibility in American options, real options approach allows the selection of optimal time point for exercise. The traditional evaluation techniques do not offer such flexible decision-making.

Given the high uncertainty in the high-tech industry, investors or management not only need to consider the R&D and manufacturing capabilities of the business, but also the impact the product will be subjected to in the market. Amram and Kulatilaka (1999) suggested that management should evaluate the extent of its ability to bear the uncertainty to explain the interaction between investment opportunity and uncertainty so as to make the optimal investment decision. Figure 1 is a perfect interpretation of the relationship between the value of investment opportunity and

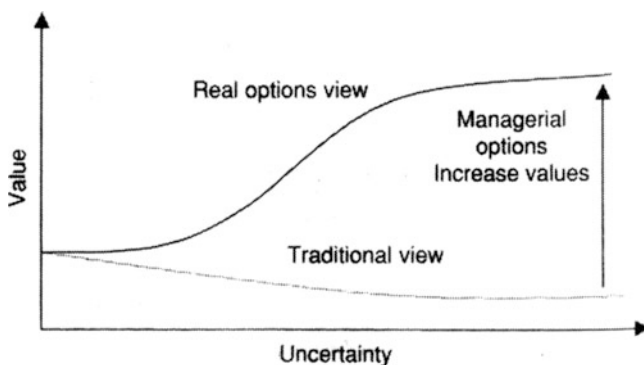


Fig. 1 Uncertainty increases value (Amram and Kulatilaka 1999)

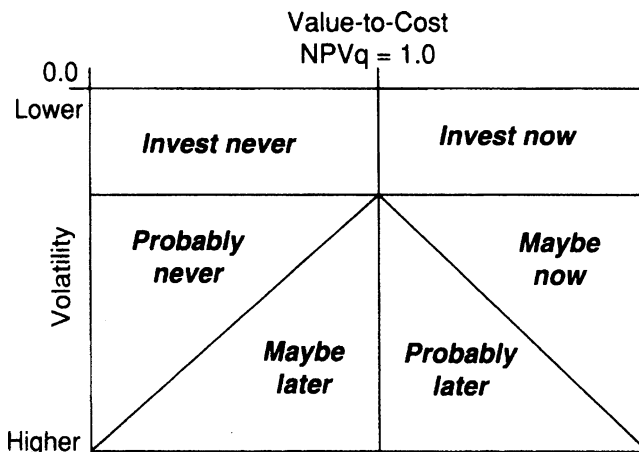


Fig. 2 Luehrman's option space

uncertainty as presented by Amram and Kulatilaka (1999); from the traditional viewpoint, an investment with high uncertainty will see its value fall. But under the viewpoint of real options, the value of an investment opportunity increases with the decisions made by the management as degree of uncertainty rises. For high-tech industry characterized by high growth, timely actions taken by the management in response to the uncertain condition can create greater value for the entire investment project. Such view is consistent with the suggestion of Trigeorgis (1996) that real options approach offers management flexibility and strategic flexibility.

Luehrman (1998a, b) used option space³ created by two option-value metrics of value-to-cost (NPV_q) and volatility⁴ to illustrate the technique of real options and locate the investment opportunity in the space for decision-making. Setting $NPV_q = 1$ as the center of abscissa, Luehrman (1998a, b) divided the option space into six regions as shown in Fig. 2, each representing a different "level" of investment opportunity, which are respectively invest now, maybe now, probably later, invest never, probably never, and maybe later. Such classification fully depicts the spirit of decision-making.

In the option space presented by Luehrman, the greater the NPV_q , the higher the cumulative variance ($\sigma^2 t$) and the higher the value of the investment project; if $NPV_q > 1$, the cumulative variance is small, indicating that other changes will be small in the future and the investment project may proceed right away. Alternately,

³Luehrman (1998a, b) estimated value-to-cost (NPV_q) by discounting the exercise price (X) at risk-free rate (r_f), which is defined as $PV(X) = X/(1 + r_f)^t$ and then as "modified" NPV after deducting the present value of exercise price from the value of the underlying (S). The value-to-cost (NPV_q) is the value of underlying asset divided by present value of exercise price, $NPV_q = S/PV(X)$; when $NPV_q > 1$, it represents the modified NPV being greater than zero; conversely, when $NPV_q < 1$, the modified NPV is less than zero. The cumulative variance ($\sigma^2 t$) is the variance of investment value multiplied by time.

⁴Defined as cumulative variance.

projects with $NPV_q > 1$ and small cumulative variance should not go ahead, while projects with large cumulative variance and relatively high uncertainty may be decided later after the inflow of new information. Projects with $NPV_q > 1$ should not be executed immediately, but wait for situation to clear up before making the decision. Using option-value metrics for investment decision-making captures the NPV of the project and the value of opportunity under high risk. The two-dimensional model of real options is perfectly interpreted in the option space of Luehrman (1998a, b).

29.2.1 Treatment of Nontraded Assets

The partial differential equation (PDE) for the pricing of derivative products derived under arbitrage-free argument may be applied regardless of whether the underlying asset is traded or not. Hull (1997) indicated that the Black–Scholes–Merton's PDE does not contain the variable of risk preferences, that is, it assumes that the risk attitude of the investors is irrelevant to the underlying. Hence, the use of risk-neutral evaluation method is meaningless in the evaluation of nontraded assets. In the real world, underlying assets to be valued are mostly nontraded assets that make the risk attitude of the investor an important factor. If the expected growth rate of the underlying asset is adjusted, it amounts to pricing the asset in a risk-neutral world.

Constantinides (1978) priced underlying asset with market risk in a world where the market price of risk is zero. He utilized the certainty equivalence approach to adjust the parameters in the model to effective value, that is, deducting risk premium and discounting the expected cash flows at the risk-free rate. The Constantinides model lets $\tilde{x}$ be cash flow, realized at the end of period, hence the risk-adjusted NPV given by capital asset pricing model (CAPM) under the assumption of single period is:

$$RANPV(\tilde{x}) = \frac{\tilde{x} - (\bar{r}_m - r_f)\text{cov}(\tilde{r}_m, \tilde{x})/\sigma_m^2}{1 + r_f} \quad (2)$$

where r_m , r_f , and σ_m are, respectively, market rate of return, risk-free return, and rate of market return shortfall. Under the assumption of zero market price of risk, $\tilde{x} - (\bar{r}_m - r_f)\text{cov}(\tilde{r}_m, \tilde{x})/\sigma_m^2$ depicts the expected cash flow ($\bar{x}$), which is adjusted to certainty-equivalent cash flows and discounted at the risk-free rate of return.⁵ Merton (1973) showed that the equilibrium security returns satisfy the basic CAPM relationship. The derivation process of Constantinides (1978) was based on the equilibrium model, while the traditional PDE-based pricing models admit no arbitrage framework. These two models have different processes, but derive consistent results.

In handling the risk factors of uncertainty, Cox et al. (1985) suggested the use of certainty equivalent cash flows, not risk-adjusted discount rate. Trigeorgis (1993a, b) also indicated that the contingent claim of asset can be priced in the real world of

⁵Cox, Ingersoll, and Ross (1985) used certainty equivalent approach for risk adjustment, not risk-adjusted discount rate, that is modifying the real expected growth rate to certainty equivalent growth rate by deducting discount premium from real expected growth rate.

systemic risk by substituting real growth rate with certainty equivalent rate. Certainty equivalent rate is obtained by deducting risk premium from the original growth rate of the asset. Such an approach is the same as pricing in a risk-neutral environment. The expected return of all assets in risk-neutral environment is risk-free return. But when the investor has certain risk preferences, the expected growth rate in equilibrium will differ from the original growth rate. Such risk adjustment approach amounts to discounting certainty equivalent cash flows at the risk-free rate of return, instead of adjusting the expected cash flows at the risk-adjusted rate.

29.2.2 Dividend-Like Yield

McDonald and Siegel (1984, 1985) discussed that since the rate of return derived from an option pricing equation should be consistent with capital market equilibrium, the results derived from the Black–Scholes equation are independent of and irrelevant to the consideration of capital market equilibrium and there exists a shortfall between the expected return and the required return. It is like dividend yield, i.e. only when the underlying asset does not pay any dividend and the expected return is equal to the market required equilibrium return will the Black–Scholes equation be satisfied. The presence of this shortfall derived from CAPM consists of the conclusion of Constantinides (1978). Trigeorgis (1993a, b) defined the shortfall as dividend-like yield (δ). Hence, if an investment project valued by real options model involves nontraded assets, the pricing model will contain a dividend-like yield, which differs from the pricing models for traded assets.

Real options pricing models apply mostly in cases of nontraded assets. In a perfect market, we assume the existence of twin security, which is a traded asset having equivalent risk as the nontraded asset and paying fixed dividend and satisfies CAPM. In such a case, the value of nontraded asset using PDE pricing model is determined under risk-neutral environment. Thus if there exists a twin security having the same financial risk as that for the nontraded asset, the real option can be priced.

Lin (2002) took into account the dividend-like yield in his real options model in the valuation of venture capital projects. Although his model simulated value of the project based on assumed parameters, the paper had comprehensive discussion of dividend-like yield. Thus under the assumption of perfect market, CAPM may be used for the estimation of dividend-like yield. Duan et al. (2003) proposed in case study that dividend-like yield can be estimated using cost of carry model that frees the estimation method from the restriction of perfect market assumption. In case study using real data, the dividend-like yields of the underlying as estimated by two different methods are close.

29.3 Hi-tech Value as a Call Option

Technological innovations and progression play an important role in driving the economic development. Countries around the world endeavor in technological innovation to maintain their competitiveness. According to the economic growth

theory of Schumpeter (1939), the innovation process is the core to understanding the economic growth and the innovation process can be divided into five patterns: production of new products, use of new technologies, development of new markets, acquisition of new materials, and establishment of any new organization. The innovation process is filled with uncertainties in every stage, from the research and development of product, to its testing, volume production, and successful entry into the market.

In the observation of old firms in the market, their competitors mostly come out of old firms in the field who either started their own business or joined other firms in the same industry. That is because the managers of older firms tend to reject innovation for the fear that it will accelerate the phase-out of existing products or that the existing production lines cannot be used for the manufacture of new products. So innovators have to leave the firm to start their own business in order to realize their innovative ideas. As a result, older firms lose many profit opportunities. According to the report by Bhidé (2000), 71% of successful entrepreneur cases made it through replication or revision of prior work experience, that is, the results of innovation. For example, Cisco is facing the threat of losing market share to Juniper; Microsoft publicly expressed in 1998 that its operating system was threatened by Linux; the microprocessors produced by Transmeta featuring low power consumption, excellent heat radiation, and low price are poised to threaten Intel and AMD. These competitors were mostly former employees of older firms. Their examples demonstrate that no matter how long a company has been established or how high its market share is, it might be replaced by new venture businesses with new ideas if it does not have webs of innovation.

The decision process for the development and investment of a new venture capital project until its IPO is a multistage investment process, which may be generally divided into seed stage, start up, growth stage, expansion stage, and bridge stage. Each stage has its missions and uncertainties, while one stage is linked to the next. In the example of Lucent New Ventures Group (LNVG) established in 1997, its investment process involved four stages: identification of opportunity, market qualifications, commercialization, and acquisition of value. The Nokia Group, founded in 1967, following merger divides its investment and development stages into production factors, investment, and innovation. From these traditional venture cases, the development and investment of a new venture business before its listing do not proceed in one stage, but in multiple stages.

In fact, the investment project involving a new venture business may be viewed as a sequential investment project. Majd and Pindyck (1987) indicated that an investment project usually possesses three properties: (1) the investment decisions and cash outflow are sequential; (2) it takes a period of time to build the project; and (3) there is cash inflow only after the project is completed. Such description fits the development stages of a venture capital project. Many Internet, biotechnology, semiconductor, and information technology companies illustrate the characteristics of negative cash flow in the initial stage, high reinvestment rate, and high uncertainty in future operations, but their IPO prices are higher than those of traditional firms. In the example of Amazon.com that had seen widening losses from 1996 to 1998, the company stock flew through the roof to US\$300; the market value of

Yahoo! once surpassed that of Boeing, the aircraft giant; Nokia lost US\$80 million in 1992, but the company enjoyed a net profit of US\$2.6 billion in 2000 after it formed the Nokia Venture Partners Fund in 1998 and its stock price once reached a P/E ratio of 100 in May 2000. The aforementioned firms are all typical venture businesses.

In recent years, venture capitalists turn their attention to biotechnology business. A large venture capital firm in Taiwan is seriously considering putting money in a biotech company in Gaithersburg, Maryland that develops immunotherapy. Analysis of its financial statements shows that the company has not been profitable in recent years. If the venture capital firm decides to invest or not based on the customary indicators, such as internal rate of return (IRR), P/E ratio, P/S ratio, and P/B ratio, it might miss a profit opportunity. When examining a venture business, investors should look at the value of infinite possible business opportunities. Referring only to numbers obtained from traditional analytic methods might result in missed investment opportunities with good profit potential.

A venture capital business usually does not focus on the sale of product or service and develops a multistage investment process. The input of funds at one investment stage begets the right to determine whether to invest in the next stage. Thus, the right of management to determine whether to invest at each stage is an American call option. After exercising the right, the management acquires the call option on strategy and management in the next stage that renders the entire investment process a multistage compound call option.

Myers (1984) suggested that the value of capital and R&D input in the initial stage of an investment project does not lie in the cash flows expected in future, but in the future growth opportunity. Therefore the investment process from the time a venture business is conceived to its mature stage or public listing may be expressed as a multistage compound option. Management can make pertinent flexible decision in response to market uncertainty in every stage to sidestep the risks brought about by uncertainty and accurately evaluate the execution of investment.

29.4 Two-Stage Compound Option

To understand the meaning of compound option, we first discuss the theory of two-stage compound option. The two-stage compound option was initiated and applied by Black–Scholes (1973), Cox and Ross (1976), Geske (1977), Roll (1977), and Myers (1987), among others.

Consider the constituents of a firm's capital structure are stocks (S) and bonds (B) and the firm has discount bonds outstanding with face value M and a maturity of T years, and suppose the firm plans to liquidate in T years and pay off the bonds. If the value of the firm V is less than M at the time of liquidation, the bondholders will get assets V and stockholders get nothing; and if V is greater than M , bondholders get M and stockholders receive $V - M$, where the payment to stockholders is $\max(V - M, 0)$. Hence, a call on the firm's stock is an option or a compound option.

According to Geske (1979), the compound option is written as

$$C(V, t) = f(S, t) = f(g(V, t), t) \quad (3)$$

Therefore, change of call value may be expressed as a function of changes in the value of the firm and time, and the dynamic stochastic process of V and C may be expressed as follows:

$$dV = \alpha_v V dt + \sigma_v V dz_v \quad (4)$$

and

$$dC = \alpha_c C dt + \sigma_c C dz_c \quad (5)$$

where α_v , α_c , σ_v , and σ_c are instantaneous expected rates of return on the firm (on the call) and instantaneous volatility of return on the firm (on the call), respectively. By applying Itô's lemma or Taylor's series expansion, the dynamics of the call option can be expressed as follows:

$$dC = \frac{\partial C}{\partial t} dt + \frac{\partial C}{\partial V} dV + \frac{1}{2} \frac{\partial^2 C}{\partial V^2} V^2 \sigma_v^2 dt \quad (6)$$

According to Merton, a multiposition, zero net investment having three-security hedge portfolio (H) consisting of the firm, a call, and a risk-free asset may be created by way of short sales and financing. Let n_1 be the dollars invested in the firm, n_2 the dollars invested in the call, and n_3 the dollars invested in risk-free debt. If dH is the instantaneous return to the hedge portfolio, then

$$dH = n_1 \frac{dV}{V} + n_2 \frac{dC}{C} + r_f n_3 dt \quad (7)$$

Substituting for the stochastic return on the firm and the call yield, we can eliminate the Weiner term of V , which implies $dH = 0$. Thus we can simplify the equation above into the familiar partial differential equation below:

$$\frac{\partial C}{\partial t} = r_f C - r_f V \frac{\partial C}{\partial V} - \frac{1}{2} \frac{\partial^2 C}{\partial V^2} V^2 \sigma_v^2 \quad (8)$$

On expiration date t^* , the value of call option is either zero or intrinsic value, subject to the boundary condition of

$$C_{t^*} = \max(0, S_{t^*} - K) \quad (9)$$

The boundary condition implies that the level of stock price will be determined by the level of call option on the value of the firm. In fact, we can learn from Eq. 8 that the variable that determines the value of option is V , not S . However, since stock is an

option on V , it follows a related diffusion and again its dynamics can be expressed as a function of V and t as:

$$\frac{\partial S}{\partial t} = r_f S - r_f V \frac{\partial S}{\partial V} - \frac{1}{2} \frac{\partial^2 S}{\partial V^2} V^2 \sigma_v^2 \quad (10)$$

Subject to the boundary condition of $S_T = \max(V_T - M, 0)$, the solution to Eq. 10 is the B–S equation:

$$S = V \Phi_1 \left(h_2 + \sigma_v \sqrt{T-t} \right) - M e^{-r_f(T-t)} \Phi_1(h_2) \quad (11)$$

and

$$h_2 = \frac{\ln(V/M) + (r_f - 0.5\sigma_v^2)(T-t)}{\sigma_v \sqrt{T-t}} \quad (12)$$

When the option expires, the decision to exercise or not depends on the relationship between S and K ; thus, at date $t = t^*$, the value of the firm, V^{cr} , that makes the holder of an option on the stock indifferent between exercising or not is the solution of the integral Equation $S_\tau^* - K = 0$ where $\tau = T - t^*$ and S_τ^* is given by Eq. 11. If the value of firm is less than V^{cr} , the call on the stock remains unexercised.

Based on Eqs. 8 and 10 as well as their boundary conditions, the two-stage compound option value given by Geske (1979) is:

$$\begin{aligned} C &= V \Phi_2 \left(h_1 + \sigma_v \sqrt{\tau_1}, h_2 + \sigma_v \sqrt{\tau_2}, \sqrt{\frac{\tau_1}{\tau_2}} \right) \\ &\quad - M e^{-r_f \tau_2} \Phi_2 \left(h_1, h_2, \sqrt{\frac{\tau_1}{\tau_2}} \right) \\ &= K e^{-r_f \tau_1} \Phi_1(h_1) \end{aligned} \quad (13)$$

and

$$h_1 = \frac{\ln(V/V^{cr}) + (r_f - 0.5\sigma_v^2)\tau_1}{\sigma_v \sqrt{\tau_1}} \quad (14)$$

$$h_2 = \frac{\ln(V/M) + (r_f - 0.5\sigma_v^2)\tau_2}{\sigma_v \sqrt{\tau_2}} \quad (15)$$

where $\Phi_2(\bullet, \bullet, \rho)$ stands for the bivariate cumulative normal distribution. The value V^{cr} is calculated using the following equation:

$$S_\tau - K = V \Phi_1(h_2 + \sigma_v \sqrt{\tau}) - M e^{-r_f \tau} \Phi(h_2) - K = 0 \quad (16)$$

with $\tau = T - t^*$, $\tau = t^* - t$, and $\tau = T - t$. Note that when $M = 0$ or $T = \infty$, the stockholder's option to repurchase the firm from the bondholders disappears and the formula reduces to that of B-S applied to a call written on the equity of an all-equity finance firm.

29.5 Multistage Real Compound Call Option and Dividend-Like Yield

Since Geske (1979) proposed the closed-form solution of compound options, the concepts of compound options have been widely applied to financial models. Kemna (1993) divided capital budgeting into the pioneer venture stage and the trial-run stage prior to commercial venture and used two-stage compound call options to assess the value of strategic flexibility of the investment projects. In Kemna's model, the underlying variable is a forward contract, while forward price is the projected value of the commercial venture at the conclusion of the trial-run stage. Because there is no cost of carry in the forwards pricing models, the problem of dividend-like yield of nontraded asset is avoided Trigeorgis (1993a, b).

Keeley et al. (1996) illustrated the multistage and high-risk nature of the investment projects with a three-dimensional tree, multistage compound call option model. This model adopted the algorithm for financial option valuation in the evaluation of an investment project. It assumed the expected return of a twin security is the risk-free interest rate, which ignored the fact that the investment project is a nontraded asset, and utilized numerical method to sidestep the difficulty of closed-form solution.

Lin (2002) employed the multistage compound call options to evaluate the multistage investment nature of high-tech firms. Lin (2002) took into account dividend-like yield for nontraded assets and compared a number of numerical solutions, but did not discuss the selection of the underlying variable and twin security. Duan et al. (2003) approached the choice problem of the underlying variable and twin security using case discussion. Their paper utilized the more efficient numerical method and proposed techniques for the setting and estimation of important parameters for multistage compound call options. It also depicted rather comprehensively the application of real options in the valuation of nontraded assets.

Thus, a perfect valuation model for high-tech investment project must take into consideration the dividend-like yield for nontraded assets and is able to reflect the uncertainty and multistage nature of the investment project and provide a complete parameter model and estimation method. In this section, we will develop a multistage compound call option equation for nontraded asset and further discuss the estimation of dividend-like yield.

29.5.1 Multistage Real Compound Call Option

Hull (1997) points out that the use of risk-neutral evaluation model is meaningless when the underlying is a nontraded asset. Constantinides (1978) used

certainty-equivalence approach to shift the underlying with market risk to a scenario where the market price of risk is zero. By referring to the paper of Merton (1973), he pointed out that an expected return of a security must satisfy the basic equation of CAPM:

$$\alpha_i - r_f = \lambda_m \sigma_{im} / \sigma_m = \lambda_m \rho_{im} \sigma_i \quad (17)$$

$$\lambda_m = (\alpha_m - r_f) / \sigma_m \quad (18)$$

where ρ , σ , and λ are correlation, standard deviation, and market price of risk, respectively. The i and m are i th security and market portfolios. Following the discussion above, if the price of any derivative is dependent on ζ and t , it should satisfy

$$\mu - r_f = \lambda \sigma \quad (19)$$

and

$$r_f = \alpha_i - \lambda_m \rho_{im} \sigma_i = \sigma_i - \lambda_i \sigma_i \quad (20)$$

Under a complete market, assuming that there exists a twin-traded security, S , having the same risk as the investment project, V , and an expected rate of return ($\alpha_s = r + \lambda_m \rho_{vm} \sigma_v$). There is a shortfall (δ), between the expected return and required return, i.e. $\delta = \alpha_s - \alpha_v = r_f + \lambda_m \rho_{vm} \sigma_v - \alpha_v$. This traded twin-traded-security must satisfy $(\alpha_v + \delta) - r_f = \lambda \sigma_v$. Thus, $(\alpha_v - \lambda \sigma_v) = r_f - \delta$ and the original PDE Eq. (8) for call option evaluation will be transformed into the following valuation model for real call options:

$$\begin{aligned} \frac{1}{2} \sigma_v^2 V^2 \frac{\partial^2 C_{t_{n-1}, t_n}}{\partial V^2} dt + (r_f - \delta) V \frac{\partial C_{t_{n-1}, t_n}}{\partial V} dV \\ + \frac{\partial C_{t_{n-1}, t_n}}{\partial t} dt - r_f C_{t_{n-1}, t_n} = 0 \end{aligned} \quad (21)$$

The constraint is $C_{m,m} = \max(V_m - I_m, 0)$, denoting only when the company value is not less than the pre-IPO investment, total I_m , is the implementation of the IPO plan meaningful.

We can follow the approach of Geske (1979) to obtain the closed-form solution of multistage compound real call option as shown below:

$$\begin{aligned} C_{(t_n - t_1)} = V e^{-\delta(t_n - t_1)} \Phi_{(n-1)}(\mathbf{H}; \mathbf{\Sigma}) \\ - \sum_{j=1}^{n-i} e^{-r_f \Phi_j} I_{t_{j+1}} \Phi_j(\mathbf{K}; \mathbf{\Sigma}) \end{aligned} \quad (22)$$

where

$$\begin{aligned}
 i &= 1, 2, \dots, n. \\
 \mathbf{H} &= (h_{i1}, h_{i2}, \dots, h_{i(n-i)})T \\
 K &= \mathbf{H} - \sigma_v \sqrt{\Psi}^T \\
 \Psi &= ((t_{i+1} - t_i), (t_{i+2} - t_i), \dots, (t_n - t_i))^T \\
 \Sigma &= \sqrt{\Psi_i / \Psi_j}, i < j. \\
 h_{ij} &= \frac{\ln\left(\frac{V}{V^{\text{cr}}}\right) + \left(r_f - \delta + \frac{1}{2}\sigma_v^2\right)(t_{i+1} - t_i)}{\sigma_v \sqrt{t_{i+1} - t_i}} \\
 h_{i(n-i)} &= \frac{\ln\left(\frac{V}{I_{t_n}}\right) + \left(r_f - \delta + \frac{1}{2}\sigma_v^2\right)(t_n - t_i)}{\sigma_v \sqrt{t_n - t_i}} \\
 V^{\text{cr}} &= (V_{i+1}^{\text{cr}}, V_{i+2}^{\text{cr}}, \dots, V_{n-1}^{\text{cr}})^T \\
 \exists V^{\text{cr}} \ni C_{n-i-1}^{\text{cr}} - I_{i+1} &= 0
 \end{aligned}$$

and

$$\begin{aligned}
 \Phi_n(\mathbf{H}; \Sigma) &= \Pr \left[\bigcap_{j=1}^n (x_j \leq h_j) \right] = \frac{1}{\sqrt{|\Sigma|(2\pi)^n}} \\
 &\int_{-\infty}^{h_1} \int_{-\infty}^{h_2} \dots \int_{-\infty}^{h_n} \mathbf{L} \int_{-\infty}^{h_n} e^{-(1/2)\mathbf{x}^T \Sigma^{-1} \mathbf{x}} d\mathbf{x}
 \end{aligned} \tag{23}$$

In seeking the value of compound options in Eq. 22, we would encounter two numerical methods. One is the solution for the approximation value of the multivariate normal integral, $\Phi(\bullet)$, and the other is the root-finding in nonlinear equation for critical value, V^{cr} . These two methods will be explored in the sections below.

29.5.2 Estimation of the Dividend-Like Yield

The estimation of dividend-like yield is important. In this section, we apply two methods – CAPM and cost of carry model for the estimation of dividend-like yield. In the CAPM approach, market risk premium and beta risk in the model are estimated first. But given that these risk values are hard to observe in the market, their estimations have many constraints. If the underlying asset has a forward contract or futures contract, the net cost of carrying the underlying can be used to estimate dividend-like yield. These two estimation methods are derived from different theories, but they reach consistent conclusions.

29.5.2.1 CAPM Method

Based on the CAPM, the expected return (α_s) to the twin security is shown as below:

$$\alpha_s = r_f + \beta \times (r_m - r_f) \quad (24)$$

where r_m is rate of market return. To estimate the required rate of return, we can compute the drift term through the dynamics of the logarithm value of underlying, and then we can obtain the drift, α_v . Andersson (1999) employed this approach to obtain the dividend-like yield and then applied it for evaluating a pulp industry with real options.

29.5.2.2 Cost of Carry Model

The reason why we estimate the dividend-like yield with the other method is because the risk premium is hard to observe in the market, and we can check how well the resulting δ is estimated by CAPM. To overcome the restriction of the perfect market assumption in CAPM, Pickles and Smith (1993) applied the formal of the future price to calculate the dividend-like yield. Their method was based on the equivalence of the expected return for the holders of an inventory and the holders of products with potential growth possibilities. The expected return for the holder of a growth product is:

$$\sigma_s = \alpha_v + \delta \quad (25)$$

where δ also equals the payout rate representing the opportunity cost of keeping the option to construct the new chipmakers. From a holding inventory point of view, the expected return is:

$$\alpha_s = \alpha_v - c + \gamma \quad (26)$$

where c is the storage cost of the commodity, γ the convenience yield of the commodity. From Eqs. 25 and 26, the dividend-like yield can be derived as:

$$\delta = \gamma - c \quad (27)$$

For two forward contracts with different expiration dates, their cost of carry differs. Without arbitrage opportunity, the difference, termed as CC, can be shown to be as follows:

$$CC = r_f + c - \gamma \quad (28)$$

where c is the cost of storage and γ the convenience yield. From Eqs. 27 and 28, we can obtain:

$$\delta = r_f - CC \quad (29)$$

The role of the dividend-like yield on the project represents the opportunity cost of keeping the option alive. The holder of an asset such as a stock expects to obtain a payout in the form of a dividend plus capital appreciation. Furthermore, it also implies that the expected net cash flows accruing from producing project, or diversification advantage of the project. The greater the dividend-like yield on the project, greater is the cost of holding the options, which also means that the investment project offers poor diversification advantage. If the dividend-like yield is relatively low or negative, this indicates that there is a lower or zero cost of holding options, or more diversification advantage.

29.6 Algorithms for Computing Multivariate Normal Integrals and Solving the Root of Nonlinear Model

Multidimensional integration is usually solved using numerical method. This chapter examines and compares three methods: Gauss quadrature method, Monte Carlo method, and Lattice method. The integration equation of cumulative distribution function is Eq. 23. Where $\mathbf{H} = (h_1, h_2, \dots, h_n)^T$, h_i are finite numbers, $\mathbf{x} = (x_1, x_2, \dots, x_n)$, Σ is symmetrical covariance matrix with positive definite eigenvalue. The multivariate normal integral cannot be obtained directly, but often computed by numerical method.

Monte Carlo method is frequently used in computing the value of multivariate normal integral. Genz (1992) successfully applied this method in solving multivariate normal integration. But the computer runtime of this method is relatively long. Gauss quadrature method is another frequently used method. Its computation becomes rather straightforward after Drezner (1992) improved upon it. This method can also be conveniently executed using the Fortran code. However, it needs the substitutions of numbers in the Gauss Integral Weights and Abscissae Table developed by Steen et al. (1969) to compute multivariate normal integral. Errors that might occur in the process are related to the order in the aforementioned tables used. Usually the higher the order, the less the error. However, the short point of this method is that as the dimension of the integration increases, so does the computer runtime. According to the examination results of Genz (1999) on a number of integration methods, Lattice method is more efficient in terms of runtime and error reduction when applied to computation of multivariate normal integral.

We utilized three numerical methods – Drezner-improved Gauss quadrature method (referred to as Drezner method⁶ below), Monte Carlo method, and Lattice method to compute multivariate normal integral so as to obtain the value of multi-stage, sequential compound options. We also compared the computer runtime of these three methods.

⁶Extend the use of bivariate normal integral computation by Drezner (1978).

29.6.1 Monte Carlo Method

To apply Monte Carlo method, Genz (1992) suggests converting the $(-\infty, h_i)$ in multivariate normal integration equation to $(0, 1)$ and covariance matrix x in Eq. 23 to Y to break n -variate normal integral into n products of one-dimensional normal integral. First one must multiply the covariance matrix X by diagonal matrix c to convert into y , then $x^T \Sigma^{-1} x = y^T y$, i.e. $c^T c = \Sigma^{-1}$, and Eq. 23 can be transformed into Eq. 30:

$$\Phi_n(\mathbf{H}) = (2\pi)^{-(n/2)} \int_{\mathbf{H}} e^{-(1/2)y^T y} dy \quad (30)$$

Equation 30 can be further developed into Eq. 31:

$$\begin{aligned} \Phi_n(H) = (2\pi)^{-(n/2)} & \int_{a'_1(y_1)}^{b'_1(y_1)} e^{-(y_1^2/2)} \dots \\ & \int_{a'_n(y_1, \dots, y_{n-1})}^{b'_n(y_1, \dots, y_{n-1})} e^{-(y_n^2/2)} dy \end{aligned} \quad (31)$$

where

$$a'_i(y) = \left(a_i - \sum_{j=1}^{i-1} c_{ij} y_j \right) / c_{ii}$$

and

$$b'_i(y) = \left(b_i - \sum_{j=1}^{i-1} c_{ij} y_j \right) / c_{ii}.$$

Let $y_i = \Phi^{-1}(z_i)$, $i = 1, 2, \dots, n$ then

$$z_i = \Phi(y_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} e^{-(t^2/2)} dt.$$

So

$$\frac{1}{\sqrt{2\pi}} e^{-(y_i^2/2)} dy_i = dz_i.$$

Then Eq. 31 may be transformed into the following:

$$\Phi_n(H) = \int_0^{e_1} \int_0^{e_2(z_1)} \dots \int_0^{e_n(z_1, z_2, \dots, z_{n-1})} dz \quad (32)$$

where

$$e_i(z_1, z_2, \dots, z_{i-1}) = \Phi \left[\left(b_i - \sum_{j=1}^{i-1} c_{ij} \Phi^{-1}(z_j) \right) / c_{i,i} \right]$$

Finally, let $z_i = e_i w_i$, $i = 1, \dots, n$, so $dz_i = e_i dw_i$, which may switch this integral interval to $[0, 1]$ interval. The product of n one-dimensional integrals is shown in Eq. 33:

$$\begin{aligned} \Phi_n(H) &= \int_0^1 \int_0^1 \dots \int_0^1 f(w_1, w_2, \dots, w_{n-1}) dw \\ &= e \int_0^1 e_2(w) \dots \int_0^1 e_n(w) \int_0^1 dw \end{aligned} \quad (33)$$

where

$$\begin{aligned} e_i(w) &= \Phi \left[\left(b_i - \sum_{j=1}^{i-1} c_{ij} \Phi^{-1}(e_j(w) w_j) \right) / c_{i,i} \right], \\ f(w) &= e_1(w) e_2(w) \dots e_n(w). \end{aligned}$$

To solve the one-dimensional integrals and their product in Eq. 33, first set the computation of initial value as $e_1 = \Phi(b_1/c_{1,1})$ and $f_1 = e_1$ to generate a set of evenly distributed random variables w_i from random sampling. The interval of even distribution is $(0, 1)$, where $i = 1, \dots, n-1$, then

$$y_{i-1} = \Phi^{-1}[w_{i-1}(e_{i-1})] \quad (34)$$

$$e_i = \Phi \left[\left(b_i - \sum_{j=1}^{i-1} c_{ij} \times y_j / c_{i,i} \right) \right] \quad (35)$$

$$f_i = e_i \times f_{i-1} \quad (36)$$

After obtaining $\bar{f}$, the average of f_i , repeat the aforementioned steps to obtain multiple $\bar{f}$ of different values and to acquire the approximation of multivariate normal integral computed by Monte Carlo method.

29.6.2 Drezner Method

Drezner's (1992) method is an improved Gauss quadrature method. In computing the multivariate normal integral, it first converts the interval in the equation from

$(-\infty, h_i)$ to $(0, \infty)$, then extracts appropriate order K^7 from the Gauss integral weight A_i and abscissae x_i table developed by Steen et al. (1969) using $W(x) = \exp(-x^2)$ as weight function and substitutes it in the equation to obtain multivariate normal integral.

In the n -dimensional integration Eq. (23), let Σ be the matrix of correlation coefficients corresponding to $\mathbf{x}$. Then $\Phi_n(\mathbf{H}; \Sigma)$ may be expressed as follows:

$$\Phi_n(H; \Sigma) = \Phi_{n-1}(H^i; \Sigma_1) - \Phi_n(H^{-i}; \Sigma_2) \quad (37)$$

where $\mathbf{H}^{-i} = (h_1, \dots, h_{i-1}, -h_i, h_{i+1}, \dots, h_n)$ and $\mathbf{H}^i = (h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_n)$. If all $\mathbf{H}$ are non-positive values, the multivariate normal integral may be computed with the following equation:

$$\Phi_n(\mathbf{H}; \Sigma) \cong C \sum_{i_1=1}^K \dots \sum_{i_n=1}^K A_{i_1} \dots A_{i_n} f(x_{i_1}, \dots, x_{i_n}) \quad (38)$$

where K is the order⁸ of weight A_i and abscissae x_i in the Gauss integral weight A_i and abscissae x_i table, while C and $f(\cdot)$ are as follows:

$$C^2 = \prod_{i=1}^n [2/r_{ii}] / [(2\pi)^n |\Sigma|] \quad (39)$$

$$f(\mathbf{u}_1, \dots, \mathbf{u}_n) = \exp \left\{ \mathbf{x}^T \mathbf{x} - (\mathbf{h}'' - \mathbf{x})^T \Sigma'' (\mathbf{h}'' - \mathbf{x}) \right\} \quad (40)$$

where

$$\begin{aligned} \mathbf{u}_i &= (h_i - \mathbf{x}_i) [\mathbf{r}_{ii}/2]^{1/2} \\ \Sigma^{-1} &= \{r_{ij}\} \\ h_i'' &= h_i [r_{ii}/2]^{1/2} \\ \Sigma'' &= \{r_{ij}''\} = \{r_{ij}/r_{ii}r_{jj}^{1/2}\} \end{aligned}$$

29.6.3 Lattice Method

Equation 33 is an n -dimensional integration equation. Lattice method first uses p -point rule to transform it into the following equation:

⁷Very low order will result in greater error. $K \approx 15$ is also the optimum choice from tables developed by Steen et al. (1969).

⁸The order K used in this chapter is 15 (see Appendices).

$$\Phi(H) = \frac{1}{p} \sum_{k=1}^p f\left(\left\{\frac{k}{p}a\right\}\right) + E \quad (41)$$

The above equation is Korobov filter, (Korobov 1957) where $\mathbf{a} = a_1, a_2, \dots, a_n$. Then $f(\mathbf{w})$ may be expressed as follows using Fourier expansion:

$$f(\mathbf{w}) = \sum_{\mathbf{m}} C_{\mathbf{m}} e^{2\pi i \mathbf{m} \cdot \mathbf{w}} \quad (42)$$

The coefficients being given by $C_{\mathbf{m}} = \int_0^1 dw_1 \dots \int_0^1 dw_n f(\mathbf{w}) e^{-2\pi i \mathbf{m} \cdot \mathbf{w}}$, where $\mathbf{m}$ is the vector of Fourier indices ($m_1, m_2, \dots, m_n$) each being an integer in $(-\infty, \infty)$, $C_{\mathbf{m}}$ is the Fourier coefficient corresponding to $\mathbf{m}$. The error of Eq. 41 can be expressed as:

$$E = \sum_{\mathbf{m}}' C_{\mathbf{m}} \Delta_p(\mathbf{m} \cdot \mathbf{a}) \quad (43)$$

where $\sum_{\mathbf{m}}$ denotes multiple summation over $(-\infty, \infty)$ in each Fourier index. The prime attached to the summation in $\sum_{\mathbf{m}}'$ denotes exclusion of $(0, 0, \dots, 0)$ from the summation. The quantity $\Delta_p(\mathbf{m} \cdot \mathbf{a})$ is defined by:

$$\begin{aligned} \Delta_p(\mathbf{m} \cdot \mathbf{a}) &= 1 \quad \text{if } \mathbf{m} \cdot \mathbf{a} = 0 \\ \Delta_p(\mathbf{m} \cdot \mathbf{a}) &= 0 \quad \text{otherwise} \end{aligned} \quad (44)$$

The method described above is the number theoretic method of Korobov. But since the Lattice method primarily purports to choose $\mathbf{a}$ value for minimization of error term E , Hlawka (1962) and Zaremba (1966) suggested that optimal $\mathbf{a}$ was chosen when the first Fourier coefficient contribution to Eq. 43 was minimized, that is,

$$\rho(a) = m_1^* m_2^* L m_n^* \quad (45)$$

where

$$m_i^* = \max(1, |m_i|)$$

But Lyness and Gabriel (1969) reckoned that this approach was the lack of an easily computable and realistic error estimate for Eq. 41. Subsequently Cranley and Paterson (1976) suggested the use of randomization to enhance the solution efficiency. Let $\boldsymbol{\beta}$ be a random vector with distribution $G(\boldsymbol{\beta})$ defined in domain of integration. Hence Eq. 45 may be rewritten as $\int R(f, \boldsymbol{\beta}) dG(\boldsymbol{\beta})$, in which when $R(f, \boldsymbol{\beta})$ may be construed, it may be randomized using the quadrature rule. In the randomized integrating range, Korobov filter may modify formula (41) into the following with the use of quadrature rules:

$$\begin{aligned} \Phi_n(H) &= \frac{1}{p} \sum_{k=1}^p f\left(\frac{k}{p}a + \boldsymbol{\beta}\right) + E_p(\boldsymbol{\beta}) \\ &= k_p(\boldsymbol{\beta}) + E_p(\boldsymbol{\beta}) \end{aligned} \quad (46)$$

The error term can be expressed as:

$$E_p(\beta) = \sum_m' C_m \Delta_p(m \bullet a) e^{2\pi i m \bullet \beta} \quad (47)$$

Equation 46 is used to compute integral using the p -point Lattice method of Korobov.

In the process of solving multivariate normal integrals in Eq. 22 to find the value of compound option, there exists a problem where the lower limit⁹ of interval is unknown, that is, we need to find critical value V^{cr} . We will discuss and compare three numerical methods for solving critical values – Newton–Raphson method, Dekker method, and Secant method. V^{cr} involves a nonlinear solving process that becomes more difficult and hard to converge as the dimension of multivariate integration increases. Frequently applied, Newton–Raphson method sometimes cannot meet the convergence requirements for finding the critical value V^{cr} . This is because Newton–Raphson method uses first-order derivative and divergence occurs when the first-order derivative is zero. The choice of critical value solver is important and should be prudent, especially if the equation contains partial derivative and other unknown factors, such as convergence requirements.

The coding of Newton–Raphson method is not difficult and easy to comprehend. But when integration also requires differentiation in the root-finding process, the coding will become more difficult. The zero method of Dekker (1969) can better satisfy the multivariate nature in computing multistage compound options. Brent (1971) further expanded on the Dekker method by combining inverse quadratic interpolation and bisection to solve the problem of zero one-dimensional partial derivative in Newton–Raphson method. It uses the method of repeatedly deducting bisections to result in faster convergence and find the root after the correct interval is defined. The computing speed of this method is also much faster.

As mentioned above, when an equation has both integration and differentiation, the Dekker method (1969) does not necessarily produce faster computing speed. Using the simple secant method instead of complicated functions offers accuracy and computer execution speed not inferior to those of Dekker's (1969) method and Brent's (1971) method. The difference between quadrature method and secant method is that the former entails repeatedly dividing the interval into bisections and approximates the root by keeping on switching the sign of function between two intervals. The latter entails using an approximating line to determine two-point interval and finding the root with this line. Their theories are similar.

⁹The lower limit of interval for the multivariate normal integral in the model has unknown critical value. After transformation into standard normal distribution, the critical value is transformed into the upper limit of interval.

29.7 Simulative Analysis

To compare the computation speed and error of three different numerical integration methods combined with two critical company value root-finding methods, we ran them with two computer programs, Fortran 4.0 and Matlab 5.3, and performed sensitivity analysis of important parameters, such as volatility and interest rate. Table 1 depicts the presumptions of parameter values required for the computation, where the project value, V , is the presumed present value of the current stage and the presumed investment in each stage is the present value of the current stage. To examine the effect of same total investment, but different distributions to each stage on the model, we classify the investment pattern into four modes: up-sloping, down-sloping, up, then down, and down, then up.

The market price of risk $\lambda_m = (r_m - r_f)/\sigma_m$, where r_m and σ_m denote the expected return and market volatility, respectively. According to the empirical results of Pindyck (1993) based on the data of New York Stock Exchange between 1926 and 1988, $r_m - r_f = 0.08$ and $\sigma_m = 0.2$. Thus, the market price of risk $\lambda_m = 0.4$, which is also presumed in the simulation computation stated below. It took an average of 6 years, $T = 6$, for a high-tech company defined in different ages to go from setup to IPO. Suppose these companies go through four stages in the process,¹⁰ we let the beginning of each stage being the major investment decision-making point and each stage have an equal duration of 1.5 years.

Table 1 Parameter value for the simulative analysis

Parameter	Symbol	Parameter value
Company value at time t_1	V	85
The investment of stage 2	I_{t2}	5 and 10
The investment of stage 3	I_{t3}	5, 10 and 15
The investment of stage 4	I_{t4}	5 and 15
The total pre-IPO investments	I_{t5}	30
Risk-free rate	r_f	[5% 13%]
Volatility of the company value	σ_v	[0.1 0.9]
Required rate of return	α_v	0.05
Market price of risk	λ_m	0.4
Correlation coefficient	ρ_{vm}	0.1 and 0.4
Maturity (year)	T	6

¹⁰A company typically goes through four stages from setup to IPO—seed, start up, expansion, and growth up. Thus our model incorporates these four stages, which form five decision points.

29.7.1 Comparing Numerical Methods for Multivariate Normal Integral and Critical Value

We compared three numerical methods for the multivariate normal integral in terms of speed and error. We also employed two computer programs to compare the computing speed of these three methods. When testing the Drezner method for multivariate normal integral, we applied the finding a zero method of Dekker (1969) in Matlab 5.3 to find critical value, V^{cr} ; when testing Lattice and Monte Carlo methods, we used the double-precision method in MS-Fortran 4.0 to solve critical value, V^{cr} . These three integration methods were coded in Matlab and Fortran languages, respectively.

Given the enormity of equations involved, we only present the results of single sensitivity. Both of two program languages used here provided built-in values¹¹ for direct computation of one-dimensional normal integral. But for a two-dimensional normal integration, only Fortran provides¹² built-in values while Matlab does not. Based on the fact that the speeds and errors of one-dimensional and two-dimensional integration are nearly identical, we applied the Gauss integration to further compare the speeds and errors in computing three-dimensional and four-dimensional integrals. From Table 2, one can see that the computer runtime of Monte Carlo method is markedly more than that of Drezner and Lattice methods. Moreover, when combining Drezner method for solving multivariate normal integral and Dekker's (1969) critical value solver, executed with the code provided in Matlab tool box, the runtime for computing three-dimensional critical value was on an average 60 s. On the other hand, running Lattice method for solving multivariate normal integral combined with secant method for solving critical value with Fortran took only 1 s.

In comparison, running the Monte Carlo method for solving multivariate normal integral combined with secant method for solving critical value with Fortran codes took on an average as high as 120 s to compute to the three-dimensional critical value. In the process of finding the critical value, V^{cr} , both Monte Carlo and Lattice methods had errors in control range and their resulting integrals are almost identical. But in terms of computing speed, there is a significant difference between them.

29.7.2 Critical Values, Company Values Against Investment Modes

The Appendix illustrates the critical values and the values of real call options.¹³ We find that all critical values in stage i are influenced by the planned investment outlay for stage i and stages thereafter. The higher the investment amount, the higher the critical value. But when the company goes for IPO, its value does not affect the critical values in stage 1 and stages thereafter. Thus the company value at the time of

¹¹ Matlab tool box provides NORCDF().

¹² DBNRDF(•) is MATLAB function code.

¹³ Referring to Lin (2002).

Table 2 Computing speed for critical value (V_t^{cr}) and real call values

Drezner's improved Gauss										Lattice		Monte-Carlo	
σ_v	V_4		V_3		V_2		V_1		V_2		V_1		
	Second	Flops	Second	Flops	Second	Flops	Second	Flops	Second	Second	Second	Second	
0.1	0.11	1,323	0.98	153,663	28.79	5,346,997	8.13	10,163,317	0.090	0.016	137.35	67.95	
0.2	0.00	1,487	1.21	165,478	36.47	6,006,569	8.12	11,697,490	0.093	0.018	129.74	61.77	
0.3	0.00	1,776	1.32	188,458	40.21	6,636,759	8.19	13,231,663	0.091	0.019	139.92	59.04	
0.4	0.00	1,873	1.53	211,306	45.81	7,568,574	8.19	14,765,721	0.110	0.019	132.19	58.32	
0.5	0.05	2,435	1.65	234,300	49.22	8,153,987	8.13	16,299,756	0.101	0.013	134.86	63.88	
0.6	0.05	2,676	1.81	257,362	46.96	7,794,794	8.13	17,833,791	0.102	0.015	129.07	61.53	
0.7	0.05	2,524	1.81	256,948	52.01	8,640,625	6.42	19,049,349	0.105	0.018	133.51	58.39	
0.8	0.06	2,917	2.09	291,705	51.96	8,658,105	5.55	20,111,421	0.103	0.019	140.82	59.77	
0.9	0.06	2,765	1.98	279,963	51.85	8,629,144	5.55	21,173,469	0.120	0.015	142.93	66.21	
Sum	0.38		14.38		403.28		66.41		0.915	0.152	1220.39	556.86	

Note: dividend-like yield, $\delta = r_f + \lambda_m \rho_{mx} \sigma_v - \alpha_v$, $r_f = 0.08$ and $\rho = 0.1$ (up-sloping investment mode)

IPO does not affect the investment decision made at each stage. From Figs. 3 and 4, one can see that investment mode affects the values of real call options. If the mode is up-sloping, the value of real call option is at the highest, followed by in sequence down-then-up mode, up-then-down mode, and down-sloping mode. These results are consistent in the three numerical integration methods described above. A down-sloping investment mode will result in the lowest value for the entire project, while up-sloping investment mode will give the entire project maximum value. These results indicate that down-sloping mode offers less decision flexibility than up-sloping mode and are consistent with the suggestions of Trigeorgis (1993a, b) for the use of real options in evaluating management and strategic flexibility.

29.7.3 Sensitivity Analysis

29.7.3.1 Sensitivity Analysis of Dividend-Like Yield

The dividend-like yield is the shortfall between the expected rate of return of a twin security (α_s) and the required rate of return to the company value (α_v). It is a function of risk-free rate (r_f), the correlation coefficient (ρ), the volatility, α_v , and the required rate of return. The smaller the ρ , the less correlation between the new investment project and the existing market portfolio and the greater diversification advantage the project offers; the greater the ρ , the less diversification advantage.

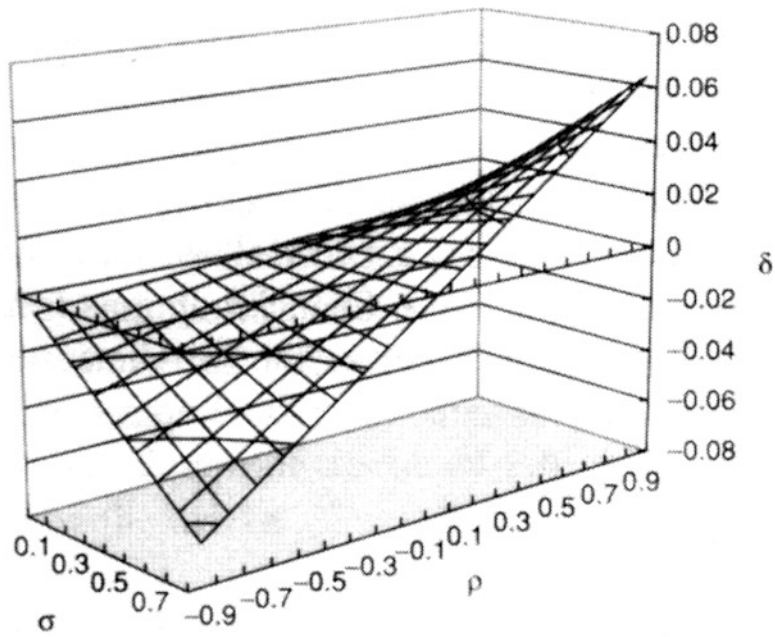
Panel A in Fig. 3 depicts the sensitivity of δ to the changes of ρ and σ . It shows that the relationship between σ and δ depends on the value of ρ ; when $\rho > 0$, σ and δ move together; when $\rho < 0$, σ increases and δ decreases. We can also show how the change of r_f and σ will affect the δ value. The results show a positive correlation between r_f and δ .

We will perform sensitivity analysis of σ and r_f , respectively, under the circumstances where ρ is equal to zero and greater than zero.

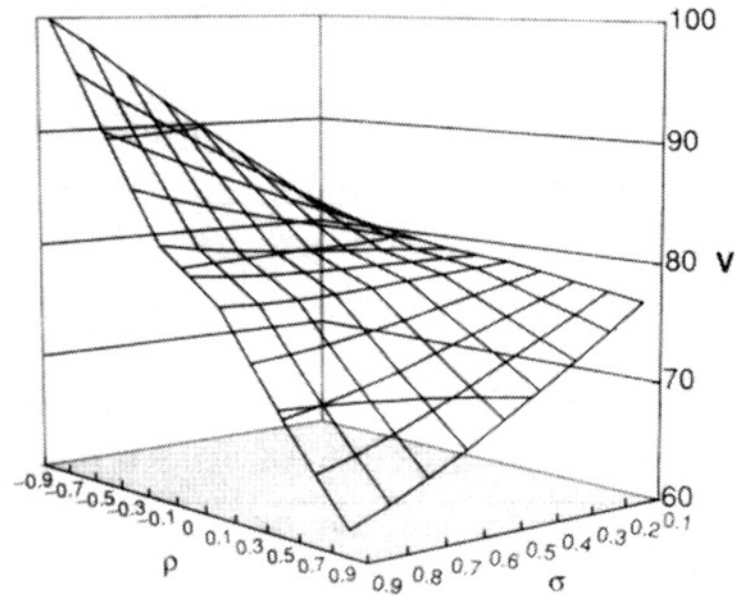
29.7.3.2 Sensitivity Analysis of Volatility

The derived closed-form solution of real call options has one difference from that of traditional financial call options. The dividend in the solution for the latter has nothing to do with volatility or interest rate, and the value of financial call option increases along with rising volatility or interest rate. But in the closed-form solution we derived for real options, dividend-like yield is an increasing function of volatility and interest rate. Therefore, real call option is not necessarily an increasing function of volatility.

The effect of correlation coefficient of up-sloping investment mode and volatility on the value of real call option under the assumption that r and α_v equal to 0.05 is shown in Panel B of Fig. 3. The results show that the greater the positive value of ρ , decreasing σ drives up the real call value, and when ρ turns negative, σ and real call value move together. It is similar to financial options framework. Fig. 4 illustrates the two-dimensional graph of σ versus value of real call option in four investment modes based on the results of three numerical integration methods under the assumption of $\rho = 0$ and 0.1, respectively.



Panel A. As $r_f = 0.05$, $\alpha_v = 0.05$



Panel B. As r_f and $\alpha_v = 0.05$

Fig. 3 Sensitivity analysis of dividend-like yield and real call values

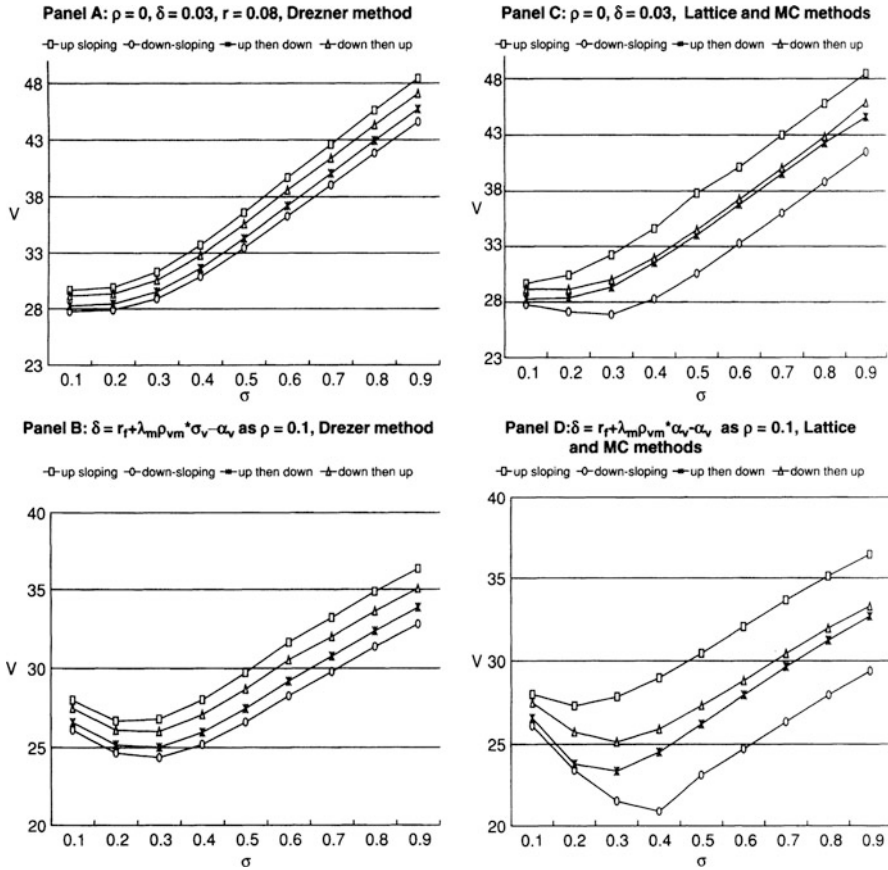


Fig. 4 Sensitivity analysis: real call option function for volatility

When $\rho = 0$, $\delta = 0.03$ and $r = 0.08$, indicating that the value of real call option increases with rising volatility. But the magnitude of increase is associated with the investment mode. Up-sloping mode showed the biggest increase, implying that decision-makers adopt an up-sloping investment approach in response to the future uncertainty of the investment project, which gives greater value on decision flexibility than the other three investment modes. This outcome is consistent in all three numerical integration methods used here. However, the integral computed by Drezner method is lower than those computed by Monte Carlo and Lattice methods under an up-sloping mode and higher under a down-sloping mode.

When $\rho > 0$, δ and σ_v move together, given that rising volatility will boost the expected rate of return to a twin security, dividend-like yield, δ , rises along with it. Under a constant r_f , the linkage between σ_v and δ becomes $\partial\delta/\partial\sigma_v = \lambda_m\rho_{vm}$.

Figure 4 shows that suppose $\rho = 0.1$, the results obtained by Drezner method show that the relationship between σ_v and real call options was akin to a J-curve. The

reflection point on the curve occurred when σ_v was 0.25, meaning real call option is a decreasing function of σ_v when $\sigma_v < 0.25$ and is an increasing function of σ_v when $\sigma_v > 0.25$. Real call option is not necessarily an increasing function of volatility. The effect of volatility on the value of investment project may not be positive. The same relationships between volatility and real call option are observed under Lattice and Monte Carlo methods that they also show a J-curve. But the value of σ_v varies at the turning point of the J-curve under different investment modes. It is approximately 0.2 under up-sloping mode and 0.4 under down-sloping mode. The result shows that the reflection point of J-curve also raises when investment mode switches from up-sloping to down-sloping. Rising volatility raises the value of option. But if we assume the investment project proceeds under a risk-aversion economic system, the required rate of return for the investors will rise with volatility, which leads to rising dividend-like yield and lowers the value of investment project. But the net effect depends on the relative magnitude of these two influences.

When the correlation coefficient of the investment project and the market portfolio gets higher, the diversification advantage the project brings to the investment portfolio is less. Figure 3 shows that after ρ rises to 0.4, σ_v has a decreasing relationship with real call options, a phenomenon totally different from the positive relationship observed between financial options and volatility. But such phenomenon becomes less apparent as risk-free rate increases. With respect to the results obtained from the Drezner method, when $\rho = 0.4$, the value of real options decreases as σ_v increases, but stays practically unchanged after $\sigma_v > 0.3$. Under both Lattice and Monte Carlo methods, the values of real call options are hardly affected by σ_v when $\sigma_v > 0.4$. It indicates that when the new project produces little diversification advantage for the market, the project is adverse to the value of the company when interest rate rises, meaning the company will have less desire to proceed with the investment.

29.7.3.3 Sensitivity Analysis of Risk-Free Rate

In the financial option model of Black–Scholes (1973), the value of call option increases when r_f rises. But in real options model, the effect of linkage between risk-free rate, r_f , and dividend-like yield (δ) must be taken into account when evaluating the effect of risk-free rate on the investment project. Dividend-like yield (δ) increases when r_f rises and the two have the following relationship:

$$\frac{\partial \delta}{\partial r_f} = 1 + \frac{\partial \lambda_m}{\partial r_f} \rho_{vm} \sigma_v \quad (48)$$

Under the circumstance where the effect of r_f fluctuation on λ_m is ignored, i.e. $\partial \delta / \partial r_f = 1$, rising r will result in drop of the project value under various values of σ_v as shown in Fig. 5 under $\rho > 0$.¹⁴ The same conclusion is observed in all three numerical integration methods discussed here. It is contrary to the theory that the

¹⁴Referring to Duan et al. (2003) under $r < 0$.

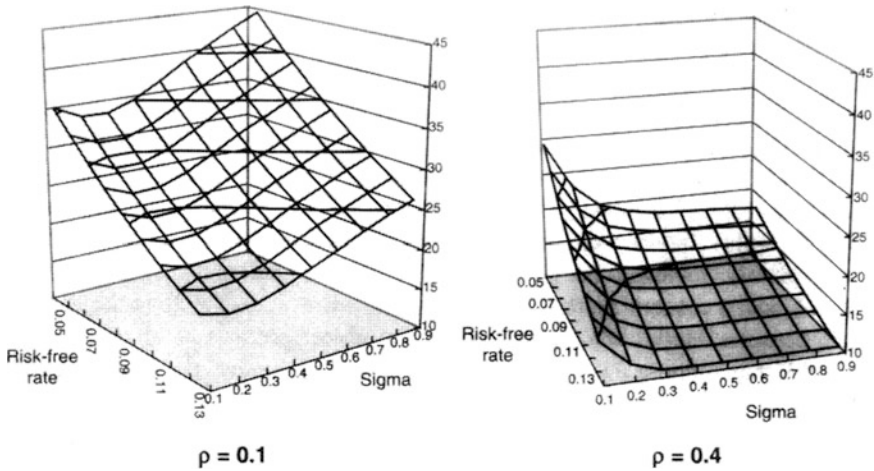


Fig. 5 Sensitivity analysis: real call option function for σ and r_f as $\rho > 0$

value of a financial call options rise with increasing r_f . Nevertheless, it is consistent with the phenomenon where higher interest rate lowers the desire of businesses to invest in new plants.

29.8 The Case Study: ProMos Technologies Inc.

ProMos Technologies Inc.¹⁵, a joint venture of Siemens Co. of Germany and Mosel Vitelic Inc., was founded in 1996 as the first plant to manufacture new generation 64 Megabytes (Mb) DRAM products in Taiwan. Today DRAM is the largest single product in the semiconductor industry and as modern technology continues to advance, more electronic products will use higher-density DRAM. For these reasons, we are interested in the value analysis of ProMos Technologies Inc.

The ProMos prospectus shows that the company mainly attempts to produce 8-in. wafer of 64 Mb DRAM chips manufactured on 0.35–0.2 μm . However, the technology¹⁶ and good ratio¹⁷ are strongly correlated to the number of chips manufactured. By this correlation, ProMos projected its monthly production to be 1550 pieces in 1997 and to reach 17,300 pieces per month by 1998, and 22,000 pieces by the time its IPO is announced.

¹⁵The ProMos expect total investment capital of US\$1,700 million in 3 years.

¹⁶The technology of manufacturing 8-in. wafer is, when manufactured on 0.35 mm, the chip ratio is 50%; when manufactured on 0.25 mm, and the chip ratio is 100%.

¹⁷The good ratio for 64 Mb DRAM is 20% at the start of mass production (1997), and it takes 1 year of production time for the good ratio to reach 80%.

29.8.1 The Model

With the IPO prospectus of ProMos, it took 3 years to complete the investment project – the foundry construction started in 1996 and the IPO was launched in 1999. We assume that there were four investment stages, with an investment decision being made at the beginning of each stage. The dynamics of underlying variable spot ASP (V_t) of DRAM follows a geometric Brownian motion. The unit production cost (I_t) is set to be the exercise price at time t , where $t = 1996, 1997, 1998$, and 1999. Hence the decision-making point for the sequential capital investments can be viewed as options on options shown in Fig. 6.

29.8.2 Finding the Underlying Variable and Twin Securities

We cannot construct a synthetic real option using the DRAM foundry because the foundry is a non-traded asset. In the DRAM markets, as the integration density of DRAM increases, the minimum feature size of DRAM decreases. Therefore, many novel process technologies have been developed in order to overcome the limitations originating in small feature-sized device and to meet the better performances in new DRAM. The competitiveness of the DRAM market drives chipmakers to keep developing higher-density chips. The density of DRAM has been approximately quadrupled every 3 years by virtue of advances of DRAM technology. Therefore, a key factor that affects the profitability of a firm depends on the successful launch of a

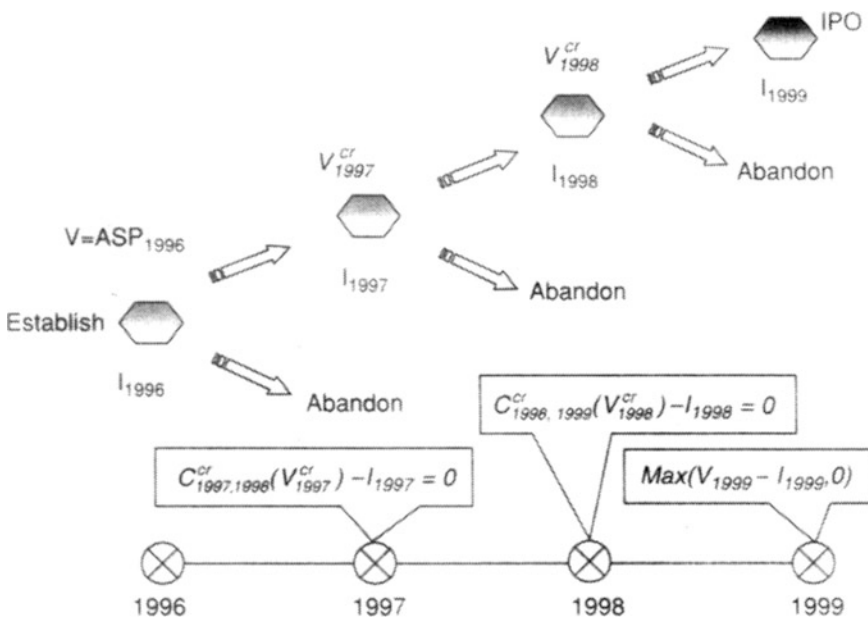


Fig. 6 Decision-making point of ProMos construction started in 1996 with the IPO in 1999

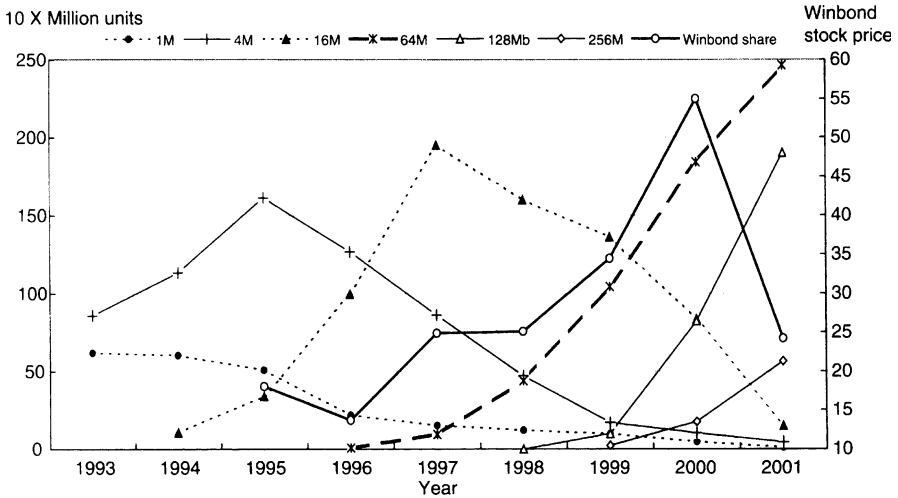


Fig. 7 Winbond Electronics Corp. share price versus quantity of sales for any DRAM product

new generation product. Figure 7 illustrates the relationship between the number of units of DRAM chips of various densities sold and the stock price for Winbond Electronics Corporation from 1993 to 2001. Figure 7 also shows that the 64 Mb DRAM products was the major product to decide the business performance of DRAM chipmakers from 1996 to 2000. The chipmaker's profitability depends mainly on DRAM sale prices and chipmaker's revenues. However, as the prospects of the semiconductor industry is closely linked to the prosperity of the DRAM industry, so a bullish electronic industry is reflected by a prospering semiconductor industry, resulting in the growth in value of prospective stock prices of the DRAM chipmakers. Therefore, if the change of stock price would influence DRAM-average sales price, the consumption goods¹⁸ can replace the twin security of investment goods.¹⁹

The common use of vector autoregressions (VARs) has been in testing for the causality between variables. Causality in the sense defined by Granger (1969) and Sims (1980) inferred that when lagged values of an independent variable have explanatory power in a regression of a dependent variable on lagged values of the dependent variable and the independent variables, the VAR can be used to test the hypothesis. Tests of the restrictions can be based on simple *F*-test in the single equation of the VAR model. The fact that the unrestricted equations have identical regressions means that these tests can be based on the results of simple ordinary least square (OLS) estimates. In Table 3, we show the causality test between the stock

¹⁸It can be DRAM product.

¹⁹It can be stock.

Table 3 Causality test between twin securities (stock portfolio) and DRAM ASP with VAR model

Variable	Model 1 ASP-R	Model 2 R_p
Intercept	-0.0034 (-0.2646)	0.0344 (1.4041)*
$ASP - R_{(-1)}$	0.4201 (3.0507)***	0.3391 (1.2994)
$ASP - R_{(-2)}$	0.4161 (2.9406)***	-0.1449 (-0.5407)
$ASP - R_{(-3)}$	-0.4034 (-2.8884)***	-0.1307 (-0.4939)
$ASP - R_{(-4)}$	-0.0822 (-0.5933)	0.3252 (1.2387)
$R_{P(-1)}$	0.0046 (0.0601)	0.1460 (1.0052)
$R_{P(-2)}$	0.0643 (0.8552)	0.2074 (1.4550)
$R_{P(-3)}$	0.0474 (0.6259)	-0.0932 (-0.6501)
$R_{P(-4)}$	-0.1643 (-2.1241)**	-0.1112 (-0.7584)
R-squared	0.4581	0.1329
F-statistic	4.8606***	0.8813

1. $ASP-R$ and R_p are the rate of return for DRAM ASP and stock portfolio, respectively; in which the lag term is indicated in the brackets

2. The rate of return for stock portfolio includes the semiconductor manufacture firms announced by the Taiwan Semiconductor Industry Association (TSIA) and which are already established in 1996. They include United Microelectronics Corp., Orient Semiconductor Electronics, Ltd., Taiwan Semiconductor Manufacturing Co., Ltd., Macronix International Co., Ltd., Mosel Vitelic Inc. and Winbond Electronics Corp.

3. ***, **, * Significant at the 0.01, 0.05, and 0.1 levels, respectively

portfolio return (R_p)²⁰ of the old DRAM chipmakers²¹ and return of DRAM ASP (ASP-R) by the VAR model, where the estimation formula of stock portfolio return is:

²⁰The weight of portfolio return is the ratio between the market value of individual stocks of six listed semiconductor firms and the total market value of the portfolio. The six listed semiconductor firms include United Microelectronics Corp., Orient Semiconductor Electronics, Ltd., Taiwan Semiconductor Manufacturing Co., Ltd., Macronix International Co., Ltd., Mosel Vitelic Inc. and Winbond Electronics Corp.

²¹The sampling firms are the semiconductor manufacture firms published by The Taiwan Semiconductor Industry Association (TSIA) since 1996.

$$R_p = \sum w_i \times R_i \quad (49)$$

$$w_i = \frac{MV_i}{\sum MV_i} \quad (50)$$

where w_i is weight and MV_i is the market capitalization of a DRAM public company i . Models 1 and 2 in Table 1 are the VAR models derived from the stock portfolio return and DRAM ASP return as independent variables, respectively. Tests reveal that only Model 1 is statistically significant (F -statistic is 4.8606) and R -squared is 0.4581. It indicates that the stock portfolio return of the existing DRAM chipmakers is the cause of DRAM ASP return and that DRAM ASP return is the result of stock portfolio return.

In 1996 ProMOS's main product was 64 Mb DRAM. As each density DRAM had a limited life cycle and in a world of rapid growth of the total bit, the DRAM price was greatly affected by cycles of new generation product. The 64 Mb DRAM was eventually overtaken by the products with higher-density levels. Therefore, the underlying of the new DRAM foundry investment project of ProMos can be used as the worldwide ASP of the DRAM market. Kelly (1998) takes the spot gold price as the underlying variable for a gold-mining investment project. We applied that the worldwide ASP as the underlying asset is different from Kelly's (1998) approach, since gold has no life cycle. The unit production cost of 64 Mb DRAM is set to be the exercise price. The options will be exercised if the spot price of DRAM goes higher than the predetermined exercise price, the new investment project is deep in the money.

29.8.3 Exercise Price

The unit cost changes over years had being set as the exercise prices shown in Table 4. The table shows that the unit cost decreases over the years since the DRAM chips were manufactured on improved technology, with 0.35 μm technology down to 0.2 μm technology from 1996 to 1999.

29.8.4 Dividend-Like Yield

Based on the CAPM, the expected return (α_s) to the twin security,²² which is a portfolio of DRAM manufacturing firms publicly listed on Taiwan stock markets, is shown as Eq. 24. According to the yearbook of Ibbotson and Sinquefeld (1999), the risk premium of market is 8%. It is similar to the conclusion of the empirical results

²²The rate of return on twin security can be expressed as $\frac{1}{4}rf + lrvmsv$, where l and rv are the market price of risk and the correlation between the underlying asset and on the twin security, respectively.

Table 4 Case study of ProMos

Panel A. Estimation of average unit production estimator	Cost symbol	1996	1997	1998	1999
Total cost ($1000 \times \text{NT\$}$) ^(a)	—	—	253,683	6,796,358	10,676,556
64 Mb DRAM manufactured on (μm)	—	—	0.35	0.25	0.25–0.2
Production technology (DRAM unit/chip) ^(b)	—	—	30	175	406
Production (pieces of wafer) ^(c)	—	—	18,500	207,168	264,000
Production (1,000DRAM units) ^{(d) = (b) \times (c)}	—	—	571	36,292	107,184
Unit cost (NT\\$) ^{(e) = (a)/(d)}	—	—	444.28	187.27	99.61
Exchange rate (NT\$/US\$) ^(f)	—	27.46	28.69	33.46	32.27
Units cost (US\$) ^{(g) = (e)/(f)}	—	—	15.480	5.597	3.08
Worldwide ASP (US\$)	—	9.690	5.909	4.101	6.415
64 Mb DRAM ASP (US\$)	—	99.42	45.82	10.19	6.59
Panel B. Estimation of the parameter					
Unit cost (exercise price)	I_t	—	15.48	5.597	3.08
Average of annual exchange rate	e_t	—	28.69	33.46	32.27
Risk-free interest rate	r_{ft}	—	5.78%	5.72%	4.14%
Volatility of the underlying	σ_v	—		37.36%	
Dividend-like yield	δ_1	—		−0.38	
	δ_2	—		−0.37203	
Worldwide DRAM ASP at 1996	V	—		9.69	
Risk-free interest rate at 1996	r_f	—		5.1%	

1. The δ_1 and δ_2 estimated by the CAPM approach and the cost of carry method respectively
2. The 64 Mb DRAM ASP is US\$99.42 in 1996

in Pindyck (1993) based on the data of New York Stock Exchange between 1926 and 1988. The beta is estimated to be -0.5485 against the market of Taiwan Stock Exchange (TSE). The expected return of the twin security is derived to be 0.797% , with a given the annual average risk-free rate of 5.185% from 1996 to 1999. Therefore, we can compute the dividend-like yield to be -0.37203 .

In the market of DRAM exchange, sales price of DRAM consists of the contract and the spot price. The pricing is the result of a negotiation between the buyer and seller. No agreement or negotiation is carried out in the trade of the spot market. Since the market trade manners differ, the response in supply–demand relationship of the two kinds of market also varies. In the buyer market, the contract price will be higher than the spot price; and the vice versa is true in the seller market. Further, due to the trade cycle of the DRAM product, the response of spot price to supply–demand is rather flexible, and therefore it would take the lead in contract price fluctuations. However, the spread differences between spot price and contract price can respond to the elasticity of supply–demand and the trading cost in the two types of trading markets.

The DRAM market is usually a buyer's market, where contract price is higher than spot price, and firms complete the trading in the contract market. Therefore, most operation revenue comes from the contract market. It is sufficient to perceive the importance of DRAM contract market in a firm's cost of carry (CC). According to a report provided by DRAMeXchange Co., the average annual percentage difference between contract price and spot price is about 10–40%. ProMos data also show that the average annual difference between contract price and spot price is approximately 20%. The study conducted by ICIS–LOR in 1999 reports that the average contract price of DRAM in the Asian market in September 1999 was US\$7.501, while the DRAM spot ASP of the same period was US\$6.769. Thus the CC amounts to 10.8%. Since an average maturity of contract takes 3 months, the annual cost of carry is 43.2%, which is also consistent with the survey of DRAMeXchange. With a risk-free interest rate of 4.41%,²³ the dividend-like yield in Eq. 29 is computed as -0.3906 . This result differs not much from the dividend-like yield obtained by the mentioned CAPM approach. The difference between the dividend-like yields estimated by those two methods is 1.857%, and therefore the computed real option is almost same.

It is important to identify whether the dividend-like yield is positive or negative. The expected return of a newly constructed DRAM foundry is obtained through the CAPM. Since the DRAM chips manufactured by old generation of technology do not have a pricing advantage, their sales prices of DRAM products always decline and the relative costs of investment opportunity will also rise. On the contrary, the price of a new generation DRAM chips is expected to rise. Therefore, the sales prices of the two different generation DRAM chips experience a negative correlation²⁴, resulting in a negative value of beta. The resulting negative value of beta implies that the diversification advantage of the newly constructed foundry contributes to the company. If the volatility is high and the risk-free interest rate is low, then a new DRAM chipmaker manufacturing higher-density DRAM with marketing strength tends to have a higher expected return and a negative dividend-like yield would occur.

29.8.5 Volatility

The worldwide ASP of the DRAM market is the basis for the new DRAM foundry investment project of ProMos. Therefore, we also employ worldwide ASP of DRAM to estimate volatility in order to avoid the calculation from being affected by the DRAM life cycle. Kelly (1998) takes the spot gold price as the underlying variable for a gold-mining investment project. We applied that the worldwide ASP as the underlying asset is different from Kelly's (1998) approach, since gold has no

²³For risk-free interest rate, we employ Government-bond average interest rate of 4.14% in 1999.

²⁴The correlation can be computed by $\text{rvm} \frac{1}{4} \beta_{\text{a}} \text{ (sm/sv)}$, where sv and sm are volatility of underlying asset and market portfolio, respectively.

life cycle. Based on the ASP calculated by Dataquest of Gartner for the worldwide DRAM spot market from 1990 to 1997, the volatility of W.W. ASP was computed to be 37.36%. These parameter estimation results are shown in Table 4.

29.8.6 Valuation of ProMos

ProMos was constructed in 1996 and IPO was made in 1999. We can obtain two critical values of V_{1997}^{cr} and V_{1998}^{cr} . If we take the worldwide ASP of DRAM in 1996 as its underlying asset, we can obtain the intrinsic value of IPO for the new DRAM foundry when it is constructed. In Table 5, we use two methods to estimate the dividend-like yield. We also place the worldwide DRAM ASP in 1996 into the model to solve for the unit value of DRAM manufactured by the DRAM chipmaker.

We find that all critical values are influenced by the unit producing cost of ProMos for stage t and stages thereafter. The higher the unit producing cost, the greater the critical value. But when ProMos goes for IPO, the W.W. ASP of DRAM does not affect the critical value. In Table 5, the V_{1998}^{cr} values are 5.83 and 5.87, respectively. It indicates that as the ASP of spot market is higher than V_{1998}^{cr} , the newly constructed DRAM foundry would exercise in 1998. Similarly, as the ASP of spot market is higher than the V_{1997}^{cr} , and it would exercise in 1997. In the table, we show that the real options values of the construction stage of the firm are US\$9.0339 and US\$8.4383, respectively. It expresses the intrinsic values of manufacturing each unit of DRAM by the firm when it is constructed.

According to the expectation of ProMos, the DRAM chips are manufactured on improved technology, 0.35–0.2 μm , and after the time of IPO, it may reach 50,000 pieces per month, as each contains 500 DRAM chips. By then, the paid-in capital of the firm may reach NT\$19.5 billion (amounts to issuing 1950 million ordinary

Table 5 IPO's value of ProMos

Method	CAPM approach	Cost of carry approach
Δ	−0.38	−0.37203
Worldwide ASP at 1996 (US\$)	9.69	
V_{1998}^{cr} at 1998	5.83	5.87
V_{1997}^{cr} at 1997	11.03	11.2
Value of real call options at 1996	9.0339	8.4383
IPO's value (US\$million)	2710.17	2531.49
IPO's value (NT\$million)	74416.47	69510.23
Price/share (NT\$)	38.16	35.65

1. The critical value V^{cr} is derived from function $C_{t_i, t_{i+1}}^{cr} \left(V_{t_i}^{cr} \right) - I_{t_i} = 0$
2. δ is dividend-like yield
3. Firm IPO's value is calculated as a value of real call options to multiply expected output
4. The share price is calculated as IPO's value divided by outstanding share when IPO

shares).²⁵ Therefore, by timing the estimated unit real options value of DRAM with the expected productivity of the firm, we can obtain the value of the firm at the time of IPO. We can also evaluate the value of shares of the firm in IPO according to the first issue of its ordinary shares. Table 5 shows the analysis of share price in IPO. It illustrates, as the underlying variable is worldwide DRAM ASP, the value per share is NT\$38.16 with $\delta = -0.38$, which is close to the listing price of NT\$40.01 of ProMos in IPO as on May 13, 1999. Since worldwide DRAM ASP already contains the price of low-density DRAM, this average would cause worldwide DRAM ASP to plunge. The value of real call options would decline. However, if the firm can upgrade its manufactured product on improved technology and actively promote the sale of higher-density DRAM in order to reduce the production cost, then the exercise price in the options model will decrease, and the product market value of the DRAM chipmaker will increase.

29.9 Conclusion

We employed the closed-form solution for multistage compound real call option of Lin (2002) to evaluate a sequential investment project. In consideration of the dividend-like yield of nontraded asset, we revised Lin's model, which did not discuss the selection of underlying variable and twin security, and referred to the method of Duan et al. (2003) for the estimation of dividend-like yield. Finally in a case study, the value of newly built DRAM foundry of ProMos was assessed.

In the process of solving multivariate normal integrals, there exists a root-finding problem for critical value at each stage where the lower limit of interval is unknown. This critical value is an at-the-money finite value. We used three numerical methods for simulating the approximation value of the multivariate normal integral and found that lattice method was the best method in terms of execution efficiency and the Monte Carlo method was the worst. But the values obtained from those two methods were close to each other. The Gauss quadrature method improved by Drezner (1992) was easy to learn and easy to apply. But it was too straightforward and ran into the selection problem at order k , resulting in critical value that differed somewhat from that obtained from the other two numerical integration methods. In the process of root-finding, we found that focus on improving the Newton–Raphson method, while ignoring the basic numerical solver tended to complicate the root-finding process.

In the simulation analysis of investment modes, it is found that the real call value of up-sloping mode was the highest among the four modes, indicating that the strategy of up-sloping investment offers more decision flexibility and provides the highest value. Such finding supports the suggestions of Trigeorgis (1993a, b) for the use of real options in evaluating management and strategic flexibility.

In volatility and interest rate sensitivity analysis, we found that the valuation result of an investment project derived from real call option differed from financial call option, mainly because an investment project a nontraded asset that is subjected

²⁵In Taiwan, the face value of per share is NT\$10.

to the influence of dividend-like yield. In the formula for dividend-like yield, there exists ρ ; when the value of ρ is large or positive, it indicates the high correlation of the new investment project with the current market status; hence poor diversification advantage of the project and higher investment risks. That is when σ increases, the value of decision flexibility brought by the new project declines, and hence the real call value drops. Conversely, when ρ is small or negative, the project offers more diversification advantage and the real call value rises with the increase of σ . Regardless of the value of ρ , when interest rate increases, the real call value drops. In addition, when ρ is large under high interest rate market, that is when the project offers poor diversification advantage, its value is subjected to the influence of interest rate only and nearly totally unrelated to σ .

In the case study, we first identified the underlying variable and twin security for the model and then applied VAR to test the causality between underlying variable and twin security. The result indicated that the use of DRAM ASP as underlying variable in the case of ProMos was appropriate. The result of dividend-like yield estimation showed that yields obtained from CAPM and cost of carry model was close, differed by only 1.857% and both values were negative, indicating the negative correlation between the new DRAM foundry of ProMos and existing portfolio of representative chipmakers. It indicates that the new investment offered diversification advantage and carried lower opportunity cost.

Based on the investment prospectus of ProMos for its new foundry that planned different process technologies in different stages, we estimated the unit manufacturing cost of DRAM and set the exercise price to find the critical value of the company and its IPO price. It is found that the at-the-money critical value decreased by the year, indicating the competitiveness of ProMos products. The manufacturing process followed the Moore's law. Finally we estimated the share price of the company at the time of IPO based on the projected capacity and its outstanding shares at the time of IPO. The result showed that the share value at the time of IPO was close to the listing price in IPO.

Acknowledgment We acknowledge the financial support from National Science Council of R.O.C.

Appendix

The firm's critical value V_i^{cr} and real call options value given δ is constant and $r_f = 0.08$

Investment mode	σ_v	V_4^{cr}	V_3^{cr}	V_2^{cr}			Real call options value		
				Drezner	Lattice	MC	Drezner	Lattice	MC
Up-sloping	0.1	43.52	51.27	52.61	52.65	52.65	29.67	29.67	29.67
	0.2	43.41	50.18	49.20	49.22	49.22	29.92	30.42	30.42
	0.3	42.77	47.85	44.45	44.50	44.50	31.30	32.25	32.25
	0.4	41.63	44.82	39.41	39.84	39.84	33.68	34.60	34.60

(continued)

				V_2^{cr}			Real call options value		
Investment mode	σ_v	V_4^{cr}	V_3^{cr}	Drezner	Lattice	MC	Drezner	Lattice	MC
	0.5	40.15	41.53	34.62	34.80	34.80	36.58	37.96	37.96
	0.6	38.48	38.25	30.30	30.47	30.47	39.70	40.14	40.14
	0.7	36.74	35.13	26.54	26.56	26.56	42.62	43.00	43.00
	0.8	35.01	32.25	23.31	23.44	23.44	45.60	45.80	45.80
	0.9	33.32	29.64	20.59	20.62	20.62	48.42	48.50	48.50
Down-sloping	0.1	32.92	41.60	55.77	55.79	55.79	27.78	27.77	27.77
	0.2	31.70	40.93	55.34	55.34	55.34	27.91	27.12	27.12
	0.3	29.76	39.14	53.60	53.72	53.72	28.90	26.87	26.87
	0.4	27.57	36.74	50.87	51.06	51.06	30.87	28.27	28.27
	0.5	25.36	34.13	47.67	47.84	47.84	33.41	30.58	30.58
	0.6	23.23	31.54	44.38	44.57	44.57	36.25	33.26	33.26
	0.7	21.26	29.08	41.21	41.47	41.47	39.04	36.00	36.00
	0.8	19.45	26.84	38.27	38.33	38.33	41.87	38.82	38.82
	0.9	17.82	24.81	35.61	35.72	35.72	44.59	41.50	41.50
Up, then down	0.1	32.92	47.09	55.06	55.13	55.13	28.28	28.28	28.28
	0.2	31.70	46.82	53.78	53.85	53.85	28.43	28.36	28.36
	0.3	29.76	45.63	51.01	51.75	51.75	29.54	29.35	29.35
	0.4	27.57	43.69	47.39	47.53	47.53	31.63	31.54	31.54
	0.5	25.36	41.38	43.51	43.76	43.76	34.28	34.03	34.03
	0.6	23.23	38.95	39.70	39.83	39.83	37.20	36.80	36.80
	0.7	21.26	36.55	36.16	36.35	36.35	40.06	39.53	39.53
	0.8	19.45	34.27	32.95	33.06	33.06	42.95	42.28	42.28
	0.9	17.82	32.16	30.11	30.13	30.13	45.72	44.59	44.59
Down, then up	0.1	43.52	45.37	53.90	53.90	53.90	29.17	29.18	29.18
	0.2	43.41	42.92	52.24	52.36	52.36	29.35	29.14	29.14
	0.3	42.77	39.47	48.94	49.06	49.06	30.56	30.00	30.00
	0.4	41.63	35.69	44.94	44.97	44.97	32.78	31.96	31.96
	0.5	40.15	31.99	40.83	40.90	40.90	35.56	34.49	34.49
	0.6	38.48	28.54	36.92	37.19	37.19	38.59	37.25	37.25
	0.7	36.74	25.44	33.38	33.58	33.58	41.41	40.06	40.06
	0.8	35.01	22.70	30.25	30.26	30.26	44.32	42.85	42.85
	0.9	33.32	20.31	27.51	27.83	27.83	47.08	45.88	45.88

References

Amram, M.H., and N.H. Kulatilaka. 1999. Real options. *Harvard Business School Press* 11: 11–31.

Andersson, H. (1999). Capital budgeting in a situation with variable utilization of capacity – an example from the pulp industry. *Working Paper, SSE/EFI*.

Bhide, A.V. 2000. *The origin and evolution of new business*. Oxford: Oxford University Press.

- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economics* 81: 637–659.
- Brennan, M.J., and E.S. Schwartz. 1985. Evaluating natural resource investment. *Journal of Business* 58: 135–157.
- Brent, R.P. 1971. *Algorithms for minimization without derivatives*. New Jersey: Prentice-Hall.
- Constantinides, G.M. 1978. Market risk adjustment in project valuation. *Journal of Finance* 33: 603–616.
- Copeland, T.E., and V. Antikarov. 2001. *Real options: A Practitioner's guide*. New York: Texere, LLC.
- Cox, J.C., and S.A. Ross. 1976. The valuation of options for alternative stochastic processes. *Journal of Financial Economics* 3: 145–166.
- Cox, J.C., J.E. Ingersoll Jr., and S.A. Ross. 1985. An inter-temporal general equilibrium model of asset prices. *Econometrica* 53: 363–384.
- Cranley, R., and T.N.L. Patterson. 1976. Randomization of number theoretic methods for multiple integration. *SIAM Journal of Numerical Analysis* 13: 904–914.
- Dekker, T.J. 1969. Finding a zero by means of successive linear interpolation. In *Constructive aspects of the fundamental theorem of algebra*, ed. Dejon and Henrici. New York: Wiley.
- Drezner, Z. 1978. Computation of the bivariate normal integral. *Mathematics of Computation* 32: 277–279.
- . 1992. Computation of the multivariate normal integral. *ACM Transactions on Mathematical Software* 18: 470–480.
- Duan, C., W.T. Lin, and C. Lee. 2003. Sequential capital budgeting as real options: The case of a new dram chipmaker in Taiwan. *Review of Pacific Basin Financial Markets and Policies* 6: 87–112.
- Genz, A. 1992. Numerical computation of the multivariate normal probabilities. *Journal of Computational and Graphical Statistics* 1: 141–150.
- . (1999). Comparison of methods for the computation of multivariate normal probabilities. *Working Paper*.
- Geske, R. 1977. The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis* 12: 541–552.
- . 1979. The valuation of compound options. *Journal of Financial Economics* 7: 63–81.
- Granger, C.W.J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–438.
- Hlawka, E.M. 1962. Zur angenäherten berechnung mehrfacher integrale. *Monatshefte Fur Mathematik* 66: 140–151.
- Hull, J.C. 1997. *Options, futures, and other derivatives*. 3rd ed. Prentice-Hall.
- Ibbotson, R.G., and R.A. Sinquefeld. 1999. *Stocks bonds, bills, and inflation yearbook*. Chicago: Ibbotson Associates.
- Keeley, R.H., S. Punjabi, and L. Turki. 1996. Valuation of early-stage ventures: Option valuation models vs. traditional approach. *Entrepreneurial and Small Business Finance* 5: 115–138.
- Kelly, S. 1998. A binomial lattice approach for valuing a mining property IPO. *The Quarterly Review of Economics and Finance* 38: 693–709.
- Kemna, A.G.Z. 1993. Case studies on real options. *Financial Management* 22: 259–270.
- Korobov, N.M. 1957. The approximate calculation of multiple integral using number-theoretic methods. *Doklady Akademmi Nauk SSSR* 115: 1062–1065.
- Lin, W.T. 2002. Computing a multivariate normal integral for valuing compound real options. *Review of Quantitative Finance and Accounting* 18: 185–209.
- Luehrman, T.A. 1998a. Strategy as Portfolio of real options. *Harvard Business Review*: 89–99.
- . 1998b. Investment opportunities as real options: Getting started on the numbers. *Harvard Business Review*: 51–67.
- Lyness, J.N., and J.R. Gabriel. 1969. Comment on a new method for the evaluation of multi-dimensional integrals. *The Journal of Chemical Physics* 50: 565–566.
- Majd, S., and R.S. Pindyck. 1987. Time to build, option value, and investment decisions. *Journal of Financial Economics* 18: 7–27.
- McDonald, R. 2002. *Derivatives markets*. Addison Wesley.

- McDonald, R., and D. Siegel. 1984. Option pricing when the underlying asset earns a below-equilibrium rate of return: A note. *Journal of Finance* 39: 261–265.
- . 1985. Investment and the valuation of firms when there is an option to shut down. *International Economic Review* 26: 331–349.
- Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867–887.
- Myers, S.C. 1977. Determinants of corporate borrowing. *Journal of Financial Economics* 5: 147–176.
- . 1984. Finance theory and financial strategy. *Interface* 14: 126–137.
- . 1987. Finance theory and financial strategy. *Midland Corporate Finance Journal* 5: 6–13.
- Pickles, E., and J.L. Smith. 1993. Petroleum property valuation: A binomial lattice implementation of option pricing theory. *Energy Journal* 14: 1–26.
- Pindyck, R.S. 1993. Investments of uncertain cost. *Journal of Financial Economics* 34: 53–76.
- Roll, R. 1977. An analytic valuation formula for unprotected American call options on stocks with known dividends. *Journal of Financial Economics* 5: 251–258.
- Schumpeter, J.A. 1939. *Business cycles: A theoretical, historical, and statistical analysis of the capitalist process*. New York: McGraw-Hill.
- Sims, Christopher A. 1980. Macroeconomics and reality. *Econometrica* 48: 1–48.
- Steen, N.M., G.D. Byrne, and E.M. Gelbard. 1969. Gaussian quadratures for integrals. *Mathematics of Computation* 23: 169–180.
- Trigeorgis, L. 1993a. The nature of option interactions and the valuation of investment with multiple real options. *Journal of Financial and Quantitative Analysis* 28: 1–20.
- . 1993b. Real options and interactions with financial flexibility. *Financial Management* 25: 202–224.
- . 1994. *Options in capital budgeting: managerial flexibility and strategy in resource allocation*. Cambridge, MA: MIT Press.
- . 1996. *Real options: managerial flexibility and strategy in resource allocation*. Cambridge, MA: MIT Press.
- Trigeorgis, L., and S.P. Mason. 1987. Valuing managerial flexibility. *Midland Corporate Journal* 5: 14–21.
- Zaremba, S.K. 1966. Good lattice points, discrepancy and numerical integration. *Annali di Matematica Pura ed Applicata* 73: 293–318.



Market Efficiency Hypothesis

30

Melody Lo

Abstract

Market efficiency is one of the most fundamental research topics in both economics and finance. Since (Fama, *Journal of Finance* 25(2): 383–417, 1970) formally introduced the concept of market efficiency, studies have been developed at length to examine issues regarding the efficiency of various financial markets. In this chapter, we review elements, which are at the heart of market efficiency literature: the statistical efficiency market models, joint hypothesis testing problem, and three categories of testing literature.

Keywords

Autocorrelation · Hypothesis testing · Information · Market efficiency · Price formation · Random walk model · Security returns · Serial correlation (tests) · (Speculative) profits · (Sub)martingale · Trading rules

30.1 Definition

The simplest but economically reasonable statement of market efficiency hypothesis is that security prices at any time fully reflect all available information to the level in which the profits made based on the information do not exceed the cost of acting on such information. The cost includes the price of acquiring the information and transaction fees. When the price formation in equity market satisfies the statement, market participants cannot earn unusual profits based on the available information. This classical market efficiency definition was formally introduced by Fama (1970), and developed at length by researchers in the field.

M. Lo (✉)

University of Texas at San Antonio, San Antonio, TX, USA

e-mail: omelody.lo@utsa.edu

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_30

937

30.2 The Efficient Market Model

Much of work on this line of research is based on an assumption that the condition of market equilibrium can be stated in terms of expected returns. Although there exists diversified expected return theories, they can in general be expressed as follows:

$$E(\widehat{p}_{i,t+1}) = [1 + E(\widehat{r}_{i,t+1}|I_t)] \times p_{i,t}, \quad (1)$$

where E is the expected value operator; $p_{i,t}$ is the price of security i in period t , $r_{i,t+1}$ is the one-period rate of return on security i in the period ending at $t + 1$, and $E(r_{i,t+1}|I_t)$ is the expected rate of return conditional on information (I) available in period t . Also, variables with hats indicate that they are random variables in period t . The market is said to be efficient, if the actual security prices are identical to their equilibrium expected values expressed in Eq. 1. In other words, if the actual security price formation follows the market efficiency hypothesis, there would be no expected returns/profits in excess of equilibrium expected returns. For a single security, this concept can be expressed as follows:

$$E(\widehat{Z}_{i,t+1}|I_t) = 0, \text{ and } Z_{i,t+1} = r_{i,t+1} - E(\widehat{r}_{i,t+1}|I_t), \quad (2)$$

where $Z_{i,t+1}$ is the return at $t + 1$ in excess of the equilibrium expected returns anticipated at t . This concept can also apply to the entire security market. Suppose that market participants use information, I_t , to allocate the amount, $\lambda_i(I_t)$, of funds available to each of n security that makes up the entire security market. If the price formation of each of n security follows Eq. 2, then the total excess market value at $t + 1$ ($\widehat{V}_{t+1}$) equals to zero, i.e.

$$E(\widehat{V}_{t+1}|I_t) = \sum_{i=1}^n \lambda_i(I_t) E(\widehat{Z}_{i,t+1}|I_t) = 0. \quad (3)$$

The general efficient market models of Eqs. 2 and 3 are the foundations for empirical work in this area. Researchers in the field largely agree that security prices “fully reflect” all available information has a direct implication: successive returns (or price changes) are independent. Consequently, researchers tend to conclude market is efficient if there are evidences that demonstrate $E(\widehat{Z}_{i,t+1}|I_t) = 0$ and $Z_{i,t}$ is uncorrelated with $Z_{i,t+k}$ for any value of k . Similarly, if $E(\widehat{V}_{t+1}|I_t) = 0$ and $V_{i,t}$ is uncorrelated with $V_{i,t+k}$ for any value of k , market is evident to be efficient.

Based on efficiency models in Eqs. 2 and 3, two special statistical models, submartingale and random walk, are closely related to the efficiency empirical literature. The market is said to follow a submartingale when the following condition holds:

$$E(\widehat{Z}_{i,t+1}|I_t) \geq 0 \text{ for all } t \text{ and } I_t. \quad (4)$$

The expected returns conditional on I_t is nonnegative and has an important implication on trading rule. This means investors should hold the security once it is bought during any future period, because selling it short cannot generate larger returns. More importantly, if Eq. 4 holds as equality, the market is said to follow a martingale. Researchers usually conclude that security prices follow “patterns” and market is inefficient when the empirical evidences are toward rejection of a martingale model.

The security prices exhibit the random walk statistical property if not only that the successive returns are independent but also that they are identically distributed. Using f to denote the density function, the random walk model can be expressed as follows:

$$f(r_{i,t+1}|I_t) = f(r_{i,t+1}) \text{ for all } t \text{ and } I_t. \quad (5)$$

The random walk property indicates that the return distributions would repeat themselves. Evidences on random walk property are often considered to be a stronger supportive of market efficiency hypothesis than those on (sub)martingale property.

30.3 The Joint Hypothesis Problem

The continuing obstacle in this line of empirical literature is that the market efficiency hypothesis per se is not testable. This is because one cannot test market efficiency hypothesis without imposing restrictions on the behavior of expected security returns. For example, the efficiency models of Eqs. 2 and 3 are derived based on a joint hypothesis: (1) the market equilibrium returns (or prices) are assumed to be some functions of the information set and (2) the available information is assumed to be fully utilized by the market participants to form equilibrium returns, and thereby current security prices. As all empirical tests of market efficiency are tests of a joint hypothesis, a rejection of the hypothesis would always lead to two possible inferences: either (1) the assumed market equilibrium model has little ability to capture the security price movements or (2) the market participants use available information inefficiently. Because the possibility that a bad equilibrium model is assumed to serve as the benchmark can never be ruled out, the precise inferences about the degree of market efficiency remains impossible to identify.

30.4 Three Categories of Testing Literature

The empirical work on market efficiency hypothesis can be categorized into three groups. First, weak-form tests are concerned with how well past security returns (and other explanatory variables) predict future returns. Second, semi-strong-form tests focus on the issue of how fast security price responds to publicly available information. Third, strong-form tests examine whether security prices fully reflect private information.

30.4.1 Weak-Form Tests

Controversy about market efficiency centers on the weak-form tests. Many results from earlier works on weak-form tests come directly from the submartingale expected return model or the random walk literature. In addition, much of the earlier works consider information set as just past historic returns (or prices). The most frequently used procedure to test the weak form of efficient markets is to examine whether there is statistically significant autocorrelation in security returns using serial correlation tests. A pattern of autocorrelation in security returns is interpreted as the possibility that market is inefficient and market participants are irrational, since they do not fully exploit speculative opportunities based on the price dependence. The serial correlation tests are tests of a linear relationship between current period's returns (R_t) and past returns (R_{t-1}):

$$R_t = \alpha_0 + \alpha_1 R_{t-1} + \varepsilon_t, \quad (6)$$

where R_t is the rate of return, usually calculated as the natural logarithm first differences of the trading price (i.e. $R_t = \ln P_t - \ln P_{t-1}$; P_t and P_{t-1} are the trading prices at the end of period t and of period $t - 1$, respectively.), α_0 is the expected return unrelated to previous returns, and α_1 is the size of first-order autocorrelation in the rate of returns. For market efficiency hypothesis to hold, α_1 needs to be statistically indifferent from 0.

After conducting serial correlation analysis, Kendall (1953) concluded that market is efficient because weekly changes in 19 indices of British industrial share prices and in spot prices for cotton and wheat exhibit the random walk property. Roberts (1959) notes that similar statistical results can be found when examining weekly changes in Dow Jones Index (see also Moore 1962; Godfrey et al. 1964; Fama 1965). Some researchers later argued that the size of serial correlation in returns offers no precise implications on the extent of speculative profits available in the market. They propose that examining the profitability of various trading rules can be a more straightforward methodology for efficiency tests. A representative study that adopted this methodology was done by Alexander (1961), where he examines the profitability of various trading rules (including the well-known $y\%$ filter rule). Despite a positive serial correlation in return series, he also discovers that $y\%$ filter rule cannot outperform buy-and-hold rule. He thus concludes that the market is still an efficient one. Similarly, Fama and Blume (1966) find positive dependence in very short-term individual stock price of the Dow Jones Industrial index. Yet, they also suggest that market is efficient because the overall trading costs from any trading rule, aiming to utilize the price dependence to profit, is sufficiently large to eliminate the possibility that it would outperform the buy-and-hold rule. In general, results from earlier work (conducted before the 1970s) provide no evidence against efficient market hypothesis since they all report that the autocorrelations in returns are very close to 0.

As more security data becomes available, the post-1970 studies always claim that there is significant (and substantial) autocorrelation in returns. Lo and MacKinlay (1988) report that there is positive autocorrelation in weekly returns on portfolios of NYSE stocks grouped according to size. In particular, the autocorrelation appears to

be stronger for portfolios of small stocks. According to Fisher's (1966) suggestion, this result could be due to the nonsynchronous trading effect. Conrad and Kaul (1988) investigate weekly returns of size-based portfolios of stocks that trade on both Wednesdays to somehow alleviate the nonsynchronous trading effect. However, as in Lo and MacKinlay (1988), they find positive autocorrelation in returns and that this pattern is stronger for portfolios of small stocks.

On another note, the post-1970 weak-form test studies focus on whether variables other than past returns can improve return predictability. Fama and French (1988) use dividend yield to forecast returns on the portfolios of NYSE stock. They find that dividend yield is helpful for return predictability. On the other hand, Campbell and Shiller (1988) report that earnings/price ratio increases the return predictability. In summary, recent studies suggest that returns are predictable when variables other than past returns are used and the evidences seem to be against the market efficiency hypothesis that was well supported before the 1970s.

30.4.2 Semi-Strong-Form Tests

Each of the semi-strong-form tests is concerned with the speed of price adjustment to a particular public information event. The event can be macro-economic announcement, companies' financial reports, or announcement on stock split. The initial work in this line of research was by Fama et al. (1969), in which they studied the speed of price adjustment to the stock-split announcement. Their results show that the informational implications of a stock split are fully reflected in the price of a share at least by the end of the month, or most probably almost immediately after the day of the stock-split announcement. They therefore conclude that the stock market is efficient because the prices respond quite speedily to new public information. Waud (1970) uses residual analysis to study how fast market reacts to the Federal Reserve Bank's announcement on discount rate changes. The result suggests that market responds rapidly to the interest-rate announcement even when the Federal Reserve Board is merely trying to bring the discount rate in line with other market rates. Ball and Brown (1968) investigate the price reactions to the annual-earnings announcement. They conclude that market participants seem to have anticipated most information by the month's end, after the annual-earnings announcement. These earlier studies (prior to the 1970s), focusing on different events of public announcement, all find supportive evidences of market efficiency hypothesis. Since the 1970s, the semi-strong-form test studies have been developed at length. The usual result is that stock price adjusts within a day of the announcement being made public. Nowadays, the notation that security markets are semi-strong-form efficient is widely accepted among researchers.

30.4.3 Strong-Form Tests

The strong-form tests are concerned with whether prices fully reflect all available information so that no particular group of investors have monopolistic access to

some information that can lead to higher expected returns than others. It is understandable that as long as some groups of investors in reality do have monopolistic access to the information, the strong-form market efficiency hypothesis is impossible to hold. In fact, both groups of specialists, NYSE (see Niederhoffer and Osborne 1966) and corporate insiders (see Scholes 1969), have monopolistic access to information, and which has been documented. Since the strong-form efficiency model is impossible to satisfy, the main focus in this line of work is to assess if private information leads to abnormal expected returns, and if some investors (with private information) perform better than others because they possess more private information. The most influential work before the 1970s was by Jensen (1968, 1969) where he assessed the performance of 115 mutual funds. Jensen (1968) finds that those mutual funds under examination on average were not able to predict security prices well enough to outperform the buy-and-hold trading rule. Further, there appears no evidence suggesting that individual mutual fund performs significantly better than what we expect from random chances. Using Sharpe–Lintner theory (see Sharpe 1964; Lintner 1965), Jensen (1969) developed a model to evaluate the performance of portfolios of risk assets. Most importantly, he manages to derive a measure of portfolio’s “efficiency.” The empirical results show that on average the resources spent by the funds managers to better forecast security prices do not generate larger portfolio returns than what could have been earned by equivalent risk portfolios selected either by random selection trading rule or by combined investments in market portfolios and government bonds. Jansen further interprets his results that probably mutual fund managers do not have access to private information. These results are clear in line with strong-form market efficiency models because evidence suggests that current security prices have fully reflected the effects of all available information. After the 1970s, there is less of new research examining investors’ access to private information that is not reflected in security prices. Representative studies were done by Henriksson (1984) and Chang and Lewellen (1984). In tests of 116 mutual funds, Henriksson (1984) reports that there is difference between mutual fund returns and Sharpe–Lintner market line. Similarly, Chang and Lewellen (1984) note that examination of mutual fund returns show no supportive evidence of fund managers’ superior selection abilities. In short, recent studies largely agree to prior literature’s view that investors with private information are unable to outperform a passive investment strategy. Evidences are still in favor of the existence of market efficiency hypothesis.

30.5 Conclusion

This review has been brief and so various issues related to market efficient model have not been considered. Volatility tests of market efficiency, and cross-sectional return predictability based on various asset pricing models are just some of the omitted issues. For more details, readers are referred to two excellent market efficiency survey papers by Fama (1970, 1991).

References

- Alexander, S.S. 1961. Price movements in speculative markets: Trends or random walks. *Industrial Management Review* 2: 7–26.
- Ball, R., and P. Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6: 159–178.
- Chang, E.C., and W.G. Lewellen. 1984. Market timing and mutual fund investment performance. *Journal of Business* 57: 57–72.
- Compbell, J.Y., and R. Shiller. 1988. Stock prices, earnings and expected dividends. *Journal of Finance* 43: 661–676.
- Conrad, J., and G. Kaul. 1988. Time-variation in expected returns. *Journal of Business* 61 (4): 409–425.
- Fama, E.F. 1965. The behavior of stock market price. *Journal of Business* 38 (1): 34–105.
- . 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25 (2): 383–417.
- . 1991. Efficient capital markets: II. *Journal of Finance* 46 (5): 1575–1617.
- Fama, E.F., and M. Blume. 1966. Filter rules and stock market trading profits. *Journal of Business (Special Supplement)* 39: 226–241.
- Fama, E.F., and K.R. French. 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22: 3–25.
- Fama, E.F., L. Fisher, M.C. Jensen, and R. Roll. 1969. The adjustment of stock prices to new information. *International Economic Review* 5: 1–21.
- Fisher, L. 1966. Some new stock-market indexes. *Journal of Business*, 39(1), Part 2: 191–225.
- Godfrey, M.D., C.W.J. Granger, and O. Morgenstern. 1964. The random walk hypothesis of stock market behavior. *Kyklos* 17: 1–30.
- Henriksson, R.T. 1984. Market timing and mutual fund performance: An empirical investigation. *Journal of Business* 57: 73–96.
- Jensen, M.C. 1968. The performance of mutual funds in the period 1945–64. *Journal of Finance* 23: 389–416.
- . 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* 42: 167–247.
- Kendall, M.G. 1953. The analysis of economic time-series, Part I: Prices. *Journal of the Royal Statistical Society*, 96 (Part I): 11–25.
- Lintner, J. 1965. Security prices, risk, and maximal gains from diversification. *Journal of Finance* 20: 587–615.
- Lo, A.W., and A.C. MacKinlay. 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1 (1): 41–66.
- Moore, A. 1962. A statistical analysis of common stock prices. PhD thesis, Graduate School of Business, University of Chicago.
- Niederhoffer, V., and M.F.M. Osborne. 1966. Market making and reversal on the stock exchange. *Journal of the American Statistical Association* 61: 897–916.
- Roberts, H.V. 1959. Stock market ‘patterns’ and financial analysis: Methodological suggestions. *Journal of Finance* 14: 1–10.
- Scholes, M. (1969). A test of the competitive hypothesis: The market for new issues and secondary offering. PhD thesis, Graduate School of Business, University of Chicago.
- Sharpe, W.F. 1964. Capital assets prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- Waud, R.N. 1970. Public interpretation of federal discount rate changes: Evidence on the ‘Announcement Effect’. *Econometrica* 38: 231–250.

The Microstructure/Micro-Finance Approach to Exchange Rates

31

Melody Lo

Abstract

The vast empirical failure of standard macro exchange rate determination models in explaining exchange rate movements motivates the development of microstructure approach to exchange rates in the 1990s. The microstructure approach of incorporating “order flow” in empirical models has gained considerable popularity in recent years, since its superior performance to macro exchange rate models in explaining exchange rate behavior. It is shown that order flow can explain about 60% of exchange rate movements versus 60% at most in standard exchange rate empirical models. As the microstructure approach to exchange rates is an active ongoing research area, this chapter briefly discusses key concepts that constitute the approach.

Keywords

Asset market approach · Currency · Divergent mappings · Exchange rates · Goods market approach · Heterogeneous information · Macroexchange rate models · Microstructure approach · Order flow · Private information · Transaction

31.1 Definition

The microstructure approach to exchange rates is considered to be a fairly new but active research area. This line of research emerged in the early 1990s mostly due to the vast empirical failure of standard macro exchange rate determination models. In more recent years (the late 1990s), there was considerably a large amount of published work regarding the microstructure approach to exchange rates, suggesting order flow is evident to be the missing piece in explaining exchange rate behavior.

M. Lo (✉)

University of Texas at San Antonio, San Antonio, TX, USA

e-mail: melody.lo@utsa.edu

The following definition of the microstructure approach to exchange rates comes directly from its pioneer, Richard Lyons (see Lyons 2001).

The microstructure approach is a new approach to exchange rates whose foundations lie in microeconomics (drawing particularly from microstructure finance). The focus of the approach is dispersed information and how information of this type is aggregated in the marketplace. By dispersed information, we mean dispersed bits of information about changing variables like money demands, risk preferences, and future inflation. Dispersed information also includes information about the actions of others (e.g. about different trading responses to commonly observed data). The fact that the private sector might be solving a problem of dispersed information is not considered in traditional macro models. Rather, macro models assume that information about variables like money demands, risk preferences, and inflation is either symmetric economy-wide, or in some models, asymmetrically assigned to a single player – the central bank. In reality, there are many types of dispersed information that exchange rates need to impound. Understanding the nature of this information problem and how it is solved is the essence of this micro-based research agenda.

31.2 Empirical Failure of Traditional Approaches to Exchange Rates

The literature has documented extensively the little ability traditional/standard exchange rate determination models have to explain exchange rate behavior. Meese and Rogoff (1983) show that a random walk model outperforms the standard international-finance models in forecasting exchange rates. In that respect, Meese (1990) writes that “. . . the proportion of (monthly or quarterly) exchange rate changes that current models can explain is essentially zero. . . This result is quite surprising, since exchange rate changes would be entirely unpredictable only in very special cases of the theoretical models discussed.” More recently, a survey paper by Frankel and Rose (1995) also notes that “To repeat a central fact of life, there is remarkably little evidence that macroeconomic variables have consistent strong effects on floating exchange rate, except during extraordinary circumstances such as hyperinflations.”

Two most frequently discussed standard exchange rate determination approaches are (1) goods market approach and (2) asset market approach. The goods market approach suggests that exchange rates move to reflect necessary changes in excess demand/supply of foreign currency resulting from international trades. A domestic economy necessarily demands for more foreign currencies when its citizens consume more imported goods. The general prediction of goods market approach is that an increase in domestic trade deficit must lead to the depreciation of domestic currency against foreign currency. However, existing studies find no empirical evidence to support any specific relation between current account imbalance and exchange rate movements.

In open economies, domestic citizens can purchase not only foreign goods but also foreign financial assets. The asset market approach suggests that demand for foreign currency increases when domestic citizens increase their possessions on foreign assets, and this in turn would cause domestic currency to depreciate against foreign currency. Different from the goods market approach, the asset market approach also concerns the market efficiency issue. Specifically, the theoretical models on asset market approach determine equilibrium exchange rate at the level that no public information can lead to excess returns.

In general, the empirical model specification for asset market approach is as follows (Lyons 2001):

$$\Delta E_t = f_1(i, m, z) + \varepsilon_{1t}, \quad (1)$$

where ΔE_t is changes in nominal exchange rate (usually monthly or weekly data is used), the function $f_1(i, m, z)$ includes the current and past values of domestic and foreign interest rates (i), money supply (m), and all other macro variables (z). Similar to the low predictability of goods market approach, the majority of asset market empirical studies report that macro variables in Eq.(1) explain 10% only, at most, of exchange rate movements. Further details on the empirical failure of various standard exchange rate determination models are well documented by Taylor (1995).

The disappointing results from the existing exchange rate models motivated researchers to look for sources responsible for the empirical failure. They attribute the general empirical failure to the unrealistic assumptions shared among standard exchange rate determination models. In detail, these models assume that every market participant learns new information at the same time when macroeconomic information/news is made public. Further, all market participants are assumed to have the ability to impound macro information into prices to the same level. However, both assumptions can easily be argued. In reality, not only market participants' information set is heterogeneous, but also their mapping ability from available information to price is impossible to be the same. The heterogeneity in information set is evident from the fact that foreign exchange traders, working for different banks, each have their own customers to deal with. Transactions with different customers offer each trader "private" information that he may not intend to share with others. In addition, it is understandable that different people tend to interpret the market impact of new information on exchange rate differently, regardless whether the information is made available to all of them at the same time. This idea of divergent mappings from information to prices is discussed by Isard (1995, pp. 182–183) who states that "economist's very limited information about the relationship between equilibrium exchange rates and macroeconomic fundamentals, ... it is hardly conceivable that rational market participants with complete information about macroeconomic fundamentals could use that information to form precise expectations about the future market-clearing level of exchange rates."

31.3 Why Microstructure Approach?

The unrealistic assumptions in standard exchange rate models mentioned above have been relaxed in the literature that aims to explain why the financial market crashed. It is important to note that despite events such as stock market crash and currency crisis appear to be macro issues, they can be largely explained by microstructure approach that considers the existence of heterogeneous information among market participants (see Grossman 1988; Romer 1993; Carrera 1999). For the same token, Lyons argues that adopting microstructure approach to investigate the trading process of exchange rates may help our understanding on when and how exchange rates move. Lyons (2001, p. 4) notes that the microstructure approach is an approach that relaxes three of the assets approach's most uncomfortable assumptions. First, on the aspect of information, microstructure models recognize that some information relevant to exchange rates is not publicly available. Second, on the aspect of players, microstructure models recognize that market participants differ in ways that affect prices. Last, on the institutional aspect, microstructure models recognize that trading mechanism differs in ways that affect prices.

31.4 The Information Role of Order Flow

The central variable that takes the fundamental role in microstructure approach, but has never been presented in any of previous exchange rate models, is order flow. Order flow is cumulative flow of signed transaction volume. A simple example on how order flow is counted for individual transaction can be helpful. Suppose that a dealer decides to sell five units of U.S. dollars via a market order (one unit usually represents a transaction worth \$1 million), then order flow is counted as -5 . The negative sign is assigned for this \$5 million transaction because it is a seller-initiated order. Each transaction is signed positively or negatively depending on whether the initiator of the transaction is buying or selling. Over time, order flow gives us a relative number of buyer-initiated versus seller-initiated orders in a market. Thus, order flow provides information to dealers about the relative demand for currencies at any time in the market. Since market participant must make buy-or-sell decisions according to available information (including their private information), it is presumed that order flow is at certain level driven by market fundamentals.

Order flow plays a fundamental role in exchange rate movements because it has the function to transmit information that is not known by everyone in the market. In fact, this concept of order flow transmitting information is intuitively appealing. As an example to describe the intuition, consider two traders (referred to dealer A and dealer B) in the foreign exchange market, and each of them trades for a particular bank. Each bank of course has its own customers from whom it buys and sells foreign exchange. When dealer A trades with his own customers, he obtains private information, such as the customers' view of the current market (price), and which, is not known to dealer B. However, when dealer A puts orders in the inter-dealer market in an attempt to balance out positions with outside customers (for inventory

concern), dealer A's private information is learned by dealer B. An alternative example is related to the idea of divergent mappings from (public) information to prices. Suppose dealer A hears a macro announcement at the same time as dealer B. Although they do not know how each other would interpret the announcement's effect on prices, they can learn this information by watching how each other trades.

A related question that is frequently asked is "does order flow really contain (market) information?" The answer is positive. The direct evidences come from dealers themselves. In surveys conducted by Cheung and Wong (2000), about 50% of dealers who responded to the survey claim that they believe banks with larger customer base have information advantage. This is because they get to trade with more customers, and more transactions ensure more private information, which leads to better speculative opportunities. Further evidence is from empirical analysis, which examine whether order flows have a permanent effect on prices. The rationale behind this empirical analysis is if order flow does not contain any information about market fundamentals, it can only have transitory effect on prices. French and Roll (1986) have used this methodology to identify the information arrival. Using vector auto-regression models, Evans (2001) and Payne (1999) found that order flow innovation has long-run effect on prices. This result provides evidence that order flow does contain information related to market fundamentals.

The general empirical model specification for microstructure approach to exchange rates can be written as follows (Lyons 2001):

$$\Delta E_t = f_2(X, I, Z) + \varepsilon_{2t}, \quad (2)$$

where ΔE_t is changes in nominal exchange rate between two transactions, function $f_2(X, I, Z)$ includes the order flow (X), dealers' inventory (I), and all other micro variables (Z). The microstructure models predict that an upward move in price is associated with a situation in which buyer-initiated trades exceed seller-initiated trades. In other words, to support microstructure approach to exchange rate, there needs to be a positive relation between order flows and prices. Lyons (2001) and Evans and Lyons (2002) have shown the considerably strong positive impact of order flow on exchange rates. More precisely, they have shown that order flow can explain about 60% (versus 10% at most in standard exchange rate empirical models expressed in Eq.(1)) of exchange rate movement.

31.5 Conclusion

The high explanatory power of order flow for exchange rate movements is exciting news for researchers in the area. So far, all empirical evidences have suggested order flow is indeed the important missing piece in exchange rate determination. Lyons (2001) thus claims that order flows help solve three exchange rate puzzles: (1) the determination puzzle, (2) the excess volatility puzzle, and (3) the forward-bias puzzle. Yet, there is not much agreement toward this claim (see Dominguez 2003). Clearly, more research needs to be done before these puzzles may be solved.

References

- Carrera, J.M. 1999. Speculative attacks to currency target zones: A market microstructure approach. *Journal of Empirical Finance* 6: 555–582.
- Cheung, Y.W., and C.Y.P. Wong. 2000. A survey market practitioners' view on exchange rate dynamics. *Journal of International Economics* 51: 401–423.
- Dominguez, K.M.E. 2003. Book review: Richard K. Lyons. The microstructure approach to exchange rates, 2001, MIT Press. *Journal of International Economics* 61: 467–471.
- Evans, M. 2001. *FX trading and exchange rate dynamics*. Working paper no. 8116, National Bureau of Economic Research.
- Evans, M., and R.K. Lyons. 2002. Order flow and exchange rate dynamics. *Journal of Political Economy* 110: 170–180.
- Frankel, J., and A. Rose. 1995. Empirical research on nominal exchange rates. In *Handbook of international economics*, vol. 3, 1689–1729. Amsterdam/New York/Oxford, UK: Elsevier/North-Holland.
- French, K.R., and R. Roll. 1986. Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics* 17: 5–26.
- Grossman, S.J. 1988. An analysis of the implications for stock and futures price volatility of program trading and dynamic hedging strategies. *Journal of Business* 61: 275–298.
- Isard, P. 1995. *Exchange rate economics*. Cambridge, UK: Cambridge University Press.
- Lyons, R.K. 2001. *The microstructure approach to exchange rates*. Cambridge, MA: MIT Press.
- Meese, R. 1990. Currency fluctuations in the post-Bretton Woods era. *Journal of Economic Perspectives* 4 (1): 117–134.
- Meese, R., and K. Rogoff. 1983. Empirical exchange rate models of the seventies. *Journal of International Economics* 14: 3–24.
- Payne, R. 1999. *Informed trade in spot foreign exchange markets: An empirical investigation*. Typescript, London School of Economics.
- Romer, D. 1993. Rational asset-price movements without news. *American Economic Review* 83: 1112–1130.
- Taylor, M.P. 1995. The economics of exchange rates. *Journal of Economic Literature* 83: 13–47.

Shashidhar Murthy

Abstract

Arbitrage is central to finance. The classical implications of the absence of arbitrage are derived in economies with no market frictions. A recent literature addresses the implications of no-arbitrage in settings with various market frictions. Examples of the latter include restrictions on short sales, different types of impediments to borrowing, and transactions costs. Much of this literature employs assumptions of continuous time and a continuous state space. This selected review of the literature on arbitrage and market frictions adopts a framework with discrete states. It illustrates and discusses a sample of the principal results previously obtained in continuous frameworks, clarifying the underlying intuition and enabling their accessibility to a wider audience.

Keywords

Arbitrage · Asset pricing · Borrowing constraints · Discrete state space · Frictions · Review · Short sales constraints · Sublinear pricing functional · Super martingales · Transactions costs

32.1 Introduction

The concept of arbitrage and the requirement that there be no arbitrage opportunities is central to finance. Essentially, an arbitrage opportunity is an investment where one can get something for nothing: a trading strategy with zero or negative current cost that is likely to yield a positive return and sure to not entail a future liability. Thus, the requirement that there be no arbitrage is a minimal desired attribute of a properly functioning securities market.

S. Murthy (✉)

Indian Institute of Management Bangalore, Bangalore, India

e-mail: shashidhar.murthy@iimb.ernet.in

The implications of the absence of arbitrage are central to much of finance, simultaneously illuminating many areas and giving rise to new fields of inquiry. From early developments of the spot-forward parity relationships to the fundamental irrelevance propositions of Modigliani and Miller (1958), many arguments have at least implicitly used the main intuition of no-arbitrage that close substitutes must obey the law of one price, viz. two securities with the same payoffs must have the same price. Modern day application of this intuition came to the fore with the Black and Scholes (1973) model of option pricing. A first systematic analysis of the implications of no arbitrage was then carried out by Ross (1976, 1978). The principal question in such analysis is: given a set of some primitive assets, how much can one infer about the valuation of other assets if there are to be no arbitrage opportunities? Both the analysis of Ross (1976, 1978) and its generalization by Harrison and Kreps (1979) assume that investors are able to trade in frictionless markets.

A recent, burgeoning literature addresses the implications of no-arbitrage in settings with various market frictions. Examples of the latter include restrictions on short sales, different types of impediments to borrowing, and transactions costs. This chapter reviews a selected portion of this literature and surveys the principal results obtained. Much of this literature employs the assumption of continuous time or an infinite dimensional state space. Here, a discrete framework is adopted in the interest of clarifying the intuition behind previously obtained results and rendering them accessible to a wider audience.

The principal implication of no-arbitrage in a frictionless setting may be summarized by what is sometimes known as the Fundamental Theorem of Asset Pricing (Dybvig and Ross 1987). This theorem states that the absence of arbitrage is equivalent to the existence of both a strictly positive linear pricing rule and a solution to the choice problem of some investor who prefers more to less. Apart from implying that the law of one price holds, this result has several alternative representations and implications. One of the best known is that the no-arbitrage value of a claim is the cost of a portfolio that exactly replicates or hedges the claim's payoff. A second is that relative prices of assets must be martingales under a "risk neutral" probability measure.

Rather than purport to be an exhaustive survey, this chapter reviews a sample of the main results from the literature on arbitrage and market frictions.¹ One striking result is that the cheapest way to hedge a given liability may be to hedge a larger liability. This was first shown by Bensaïd et al. (1992) in a transactions costs setting. An implication of this is that pricing may fail to be linear and instead be sublinear: the value of the sum of payoffs may be less than the sum of the values of the individual payoffs. Thus, there may be room for financial innovation, or departures from Modigliani–Miller (1958) type irrelevance, where an intermediary pools securities, and then strips them; see Chen (1995) for a discussion. When there are no frictions, the price paid when buying a claim is also the amount received in going short or writing the claim. Market frictions which result in sublinearity of the valuation or pricing rule can lead to bid–ask spreads on derivative securities even when there are no transactions costs (i.e. bid–ask spreads) in trading the primary securities, as shown by Luttmer (1996).

Furthermore, departures from the law of one price and the martingale property may occur under frictions. In the presence of a short sales constraint that changes elastically depending on the collateral posted. Hindy (1995) showed that an asset's value depends not only on its dividends but also on the collateral services it provides. When investors face short sales or borrowing constraints, Jouini and Kallal (1995a, b) show that asset prices may be super martingales.

The rest of this chapter is organized as follows. A basic framework is set out in the next section, following which the benchmark case of no frictions is discussed in Sect. 32.3. Due to limitations of space, we formally illustrate the above results considering primarily the case of no short sales in Sects. 32.4 and 32.5. However, we also briefly outline the impact of other types of frictions such as constraints on portfolio weights that permit some short sales (such as that under a leverage constraint or margin restriction), and transactions costs in Sect. 32.6. We conclude with some remarks relating to the consistency with equilibrium of results obtained from the no-arbitrage approach under frictions.

32.2 A Basic Framework

Consider an economy over dates $t = 0$ and T . Uncertainty is described by a discrete state space Ω with typical member $\omega \in \{1, \dots, N\}$ denoting the final state of nature realized at date T where $N < \infty$. The probabilities of these states are $\{p(\omega)\}$ corresponding to an underlying probability measure P .

Investors trade a set of primitive assets which are in positive net supply, and whose prices are taken as given. Asset $j = 1, \dots, J$ has price $S_j(0)$ at date 0 and the future price $S_j(\omega) \equiv D_j(\omega)$ in state ω at date T , where D_j is a given random dividend or payoff. Asset $j = 1$ is taken to be a risk-free bond with current price of unity; (one plus) its constant interest rate is denoted R . A portfolio choice is $z \equiv (z_1, \dots, z_J)$, comprising holdings of shares of the various assets at date 0. Investors choose portfolios to maximize their preferences that are strictly increasing in consumption at dates 0 and T .

Trading in assets is subject to market frictions that take the form of a constraint on short sales and/or borrowing. The formulation we will consider for most of this chapter restricts holdings of shares of some or all assets to be at least as large as exogenously given lower bounds: $z_j \geq -\underline{z}_j$, where $\underline{z}_j \geq 0$.² In the case of no short sales of asset j , $\underline{z}_j = 0$; if instead some limited but fixed amount of short sales is permitted, $\underline{z}_j > 0$. Similarly, note that the no borrowing case corresponds to $\underline{z}_1 = 0$, since asset $j = 1$ is the risk-free bond. A portfolio that satisfies the short sales constraint is termed admissible.

Investors can use the primitive assets to create, i.e. exactly replicate, various payoffs using admissible portfolios. Every such payoff $x \equiv \{x(\omega)\}$, where $x(\omega) = \sum_j z_j S_j(\omega)$ is hence said to be marketed, i.e. available for purchase and/or sale. In the presence of market frictions, the set of marketed payoffs is not limited to those payoffs that can be explicitly replicated. For instance, consider a payoff x of 1 in some state ω' and 0 in other states whose replication require a portfolio that involves a short position in asset j (and positions in other securities). Suppose, the latter short position is equal to the

maximum amount permitted of $z_j > 0$. Then, the payoff $2x$ cannot be exactly replicated because it would require a short position of $2z_j$ shares. However, the payoff $2x$ may still be termed marketed if there exists a portfolio that produces *at least* 2 in state ω' and 0 elsewhere; i.e. if the payoff can be *super-replicated*.

$$\phi(x) \equiv \left\{ \sum_j z_j S_j(0) : x(\omega) \leq \sum_j z_j S_j(\omega), \forall \omega \right\} \quad (1)$$

Thus, it is natural to define a price for an arbitrary payoff x as the minimum cost at which it can be exactly replicated or super-replicated by an admissible portfolio, where the associated functional $\phi(\cdot)$ is termed a pricing or valuation rule.³

An arbitrage opportunity is an admissible portfolio z that either has (1) a non-positive cost $\sum_j z_j S_j(0)$ when initiated and a date T payoff $x \equiv \{x(\omega)\}$, where $x(\omega) = \sum_j z_j S_j(\omega)$, which is positive in some states and nonnegative in others, or (2) a negative current cost and a nonnegative future payoff in all states.

32.3 Exact Replication and Prices under no Frictions

At this stage, it is useful to present the principal result on the implications of the absence of arbitrage for the benchmark case where there are no market frictions. This result, known as the Fundamental Theorem of Asset Pricing, is due to Ross (1976, 1978). Given the definition of the pricing or valuation operator $\phi(\cdot)$, it is clear that there are no arbitrage opportunities in this frictionless setting only if every nonnegative marketed payoff x (which is also positive in some state) has price $\phi(x) > 0$. The result below establishes a further property: that of linearity. See Dybvig and Ross (1987) for a proof of the result below.

Proposition 1

Suppose there are no market frictions, i.e. $z_j = \infty, \forall j$. Then there are no arbitrage opportunities if and only if the pricing rule in Eq. 1, denoted $\phi^(\cdot)$ here, is positive and linear.*

Apart from implying that the law of one price must hold, the linearity property means that $\phi^*(\lambda x) = \lambda \phi^*(x)$ for all λ , i.e. the price functional is homogeneous. It is useful to further interpret the above result in terms of an implicit state price vector $\xi \equiv \{\xi(\omega)\}$, where $\xi(\omega)$ is the price of a state security that pays 1 unit in state ω , and 0 elsewhere. The linearity and positivity of $\phi^*(\cdot)$ are equivalent to $\phi^*(x) = \sum_{\omega} \xi(\omega) x(\omega)$ and $\xi(\omega) > 0$, respectively. The pricing rule $\phi^*(\cdot)$ values every marketed payoff precisely because the latter can be exactly replicated, or hedged, using a portfolio of existing assets: it assigns a value equal to the cost of the replicating portfolio.

Another useful interpretation of the linearity of $\phi^*(\cdot)$ is that there exists a (“risk-neutral”) probability measure $\mathcal{Q}^*$ that is equivalent to the underlying measure P under which relative or normalized asset prices are martingales. Thus, every primitive asset’s current price relative to, say, the price of the bond (which is 1), is equal to the expectation under $\mathcal{Q}^*$ of its future payoffs relative to that of the bond: $S_j(0) = E^{\mathcal{Q}^*}[D_j R^{-1}]$. Equivalently, the value of every payoff satisfies $\phi(x) = \sum_{\omega} q^*(\omega) x(\omega) R^{-1}$, where $q^*(\omega)$ denotes the risk-neutral probability of state ω under $\mathcal{Q}^*$. These well known implications of no-arbitrage in frictionless markets provide the basis of most option pricing models, following Black and Scholes (1973), Merton (1973), and Cox and Ross (1976).

32.4 No Short Sales

We now return to the economy with frictions of Sect. 32.2, and consider the case of no short sales. As in the frictionless case, it is clear that there are no arbitrage opportunities in this setting only if every nonnegative marketed payoff x (which is positive in some state) has price $\phi(x) > 0$. We proceed by recording a result below that is the counterpart to Proposition 1.

Proposition 2

Suppose the only friction is that the short sales of some assets is prohibited, i.e. $z_j = 0$ for some j , and $z_j = \infty$ for the rest. Then there are no arbitrage opportunities if and only if the pricing rule in Eq. 1, denoted $\phi^{NS}(\cdot)$ here, is positive and sublinear. Furthermore, there exist underlying positive hypothetical linear pricing rules $\phi(\cdot)$ such that $\phi^{NS}(x) \geq \phi(x)$, for all marketed payoffs x . Also, there exists a new probability measure associated with $\phi^{NS}(\cdot)$ under which the (normalized) price process of an asset is a super martingale if the asset cannot be sold short, and a martingale if the asset can be sold short.⁴

The proof follows from Garman and Ohlson (1981), Chen (1995), Jouini and Kallal (1995a, b), and Luttmer (1996) and rather than reproduce it here, we will shortly present a simple binomial example where the result is explicitly illustrated. (Also note that while some of these papers consider transactions costs, their results apply here). But first, a few implications of the sublinearity property and the supermartingale property are discussed.

Observe that, in contrast to Proposition 1, the pricing rule $\phi^{NS}(\cdot)$ is not linear but sublinear. The sublinearity implies that the value of a portfolio of two payoffs x and y may be less than the sum of the values of the payoffs, i.e. $\phi^{NS}(x + y) \leq \phi^{NS}(x) + \phi^{NS}(y)$.

It also implies that $\phi^{NS}(\lambda x) = \lambda \phi^{NS}(x)$ for all $\lambda \geq 0$, i.e. the price functional is positively homogeneous.

Chen (1995) discusses the role of financial innovation in such a context. He shows that an innovator (who is assumed to not face any short sales constraint, unlike other investors) can earn profits by purchasing a “pooled” payoff $x + y$ at a cost $\phi^{NS}(x + y)$, stripping it into individual components x and y , and selling (i.e. issuing) the latter at prices $\phi^{NS}(x)$ and $\phi^{NS}(y)$, respectively. Other investors cannot earn the same profits because they cannot short-sell (i.e. issue) the individual component securities x and y . In a frictionless economy, in contrast, the linearity of the pricing rule $\phi^*(\cdot)$ leaves no role for such financial innovation; i.e. the Modigliani–Miller (1958) invariance proposition holds.

Next, consider the relationship between the value of a security with payoff x and another security with payoff $-x$. In a frictionless world, the values of these two securities (the second security is essentially equivalent to going short the first) are the negative of each other, i.e. their values sum to 0. This follows from the linearity (homogeneity) of the valuation rule ϕ^* . Under no short sales, the valuation rule $\phi^{NS}(\cdot)$ is only positively homogeneous, and thus $\phi^{NS}(-x)$ may differ from $-\phi^{NS}(x)$. The intuition is just that the cost of super-replicating a payoff x will in general differ from that for the payoff $-x$. Also note that since the value of a zero payoff must be zero, $\phi^{NS}(x) + \phi^{NS}(-x)$ is at least as large as $\phi^{NS}(x + (-x)) = \phi^{NS}(0) = 0$; i.e. the sum of the values of both securities may be positive. Consequently, the ask price $\phi^{NS}(x)$ of the payoff x may exceed the bid price $-\phi^{NS}(-x)$. Thus, as Jouini and Kallal (1995a, b) and Luttmer (1996) show, a derivative security’s price may exhibit a bid–ask spread even where there are no transactions costs (i.e. bid–ask spreads) in trading the primitive assets.

As we noted in Sect. 32.3, asset prices (normalized by, say, the bond) in frictionless economies are martingales under the risk-neutral probability measure. In other words, one cannot expect to earn more than the risk-free rate after correcting for risk. In sharp contrast, Proposition 2 shows that there exists a risk-neutral probability measure, say Q^{NS} , under which (normalized) prices of assets subject to short sales constraints are super martingales. In other words, $S_j(0)/R^{-1} \geq E^{Q^{NS}}[D_j]$ for such assets: their prices after correcting for risk and the risk-free return are expected to be nonincreasing. This is compatible with the absence of arbitrage opportunities from the perspective of a risk-neutral investor because an asset whose price is expected to decrease relative to the bond cannot be sold short. This super martingale property was proved by Jouini and Kallal (1995a, b) in a model with short sales constraints (and transactions costs).

32.5 A Simple Binomial Model

As an example of a simple model that explicitly illustrates the results of Proposition 2 and their significance, we now consider a one-period binomial model. A stock and bond are traded with the constraint that no short sales of the stock is permitted, but borrowing (short sales of the bond) is allowed. The stock’s current price is S and its end-of-period price is uS in state u , and dS in state d . The bond has current price of unity and one plus a risk-less return of R where $d < R < u$.

Consider a payoff $x \equiv (x_d, x_u)$ comprised of x_d in state d and x_u in state u . Hedging any such payoff requires a portfolio of z_s shares of the stock and z_b units of the bond that satisfies

$$z_s \omega S + z_b R \geq x_\omega \text{ and } z_s \geq 0, \quad (2)$$

where $\omega \in \{d, u\}$ denotes both the future state and the return of the stock. Note from Eq. 2 we allow for the possible super-replication of the payoff; also observe that z_s must satisfy the no-short-sales constraint. Since the cost of the hedge portfolio is $z_s S + z_b$ it follows, using Eqs. 1 and 2, that the value of the payoff is

$$\begin{aligned} \phi^{NS}(x) &\equiv \text{Min } \{z_s S + z_b : z_s \omega S + z_b R \geq x_\omega; \\ &\quad z_s \geq 0; \\ &\quad \omega \in \{d, u\}\}, \end{aligned} \quad (3)$$

i.e. it equals the cost of the cheapest hedge portfolio.

Denote the risk-neutral probability of state u in the frictionless counterpart to the above example by $q^* \equiv (R - d)/(u - d)$. It is then easy to verify that the solution to Eq. 3 is:

$$\phi^{NS}(x) = [q^* x_u + (1 - q^*) x_d] R^{-1} \text{ if } x_u \geq x_d \quad (4)$$

and

$$\phi^{NS}(x) = x_d R^{-1} \text{ if } x_u < x_d. \quad (5)$$

In other words, for a payoff such as that of a call option, where $x_u > x_d$, the value is given by Eq. 4 and is no different from what it would be in a frictionless world. This is because exact replication, or an exact hedge, of the call entails a long position in the stock and borrowing. In contrast, for a security such as a put option, where $x_u < x_d$, the value in Eq. 5 is just the discounted value of the payoff in the “down” state discounted at the risk-free return. The reason is that an exact hedge or replication of the put would require short sales of the stock and is hence infeasible due to the no-short-sales constraint. Instead, the cheapest super-replication of the put involves a long bond position with face value x_d .

To see that the valuation functional ϕ^{NS} in Eqs. 4 and 5 is sublinear, compare the value of the payoff (dS, uS) from the stock with the sum of the values $\phi^{NS}(dS, 0)$ and $\phi^{NS}(0, uS)$. The former is obviously $\phi^{NS}(dS, uS) = [q^* uS + (1 - q^*) dS] R^{-1} = S$. However, the latter sum, $\phi^{NS}(dS, 0) + \phi^{NS}(0, uS) = dSR^{-1} + [q^* uS + (1 - q^*) 0] R^{-1} = S + dSR^{-1} q^*$, exceeds the current stock price, and this proves the sub-linearity. The intuition is that the cost of hedging the combined payoff (dS, uS) is less than the sum of the costs of hedging $(dS, 0)$ and $(0, uS)$ because hedging $(dS, 0)$ entails super-replication.

Finally, we show how the super martingale property of Proposition 2 comes about. Recall that with no frictions, $q^* \equiv (R - d)/(u - d)$ is the risk-neutral probability of state u under which the stock, bond, and all other payoffs (i.e. options) are martingales. Now define the probability $q \in [0, q^*]$ and the associated hypothetical linear valuation rule $\phi^q(x) = [qx_u + (1 - q)x_d]R^{-1}$. It is easy to verify that the actual sublinear valuation rule $\phi^{NS}(\cdot)$ of the economy with short sales constraints in Eqs. 4 and 5 is related to the sets $\{q\}$ and $\{\phi^q(\cdot)\}$ by:

$$\phi^{NS}(x) = \text{Max}\{\phi^q(x) : q \in [0, q^*]\}. \quad (6)$$

Compared to the probability q^* , every other probability $q \in [0, q^*]$ places less weight on the “up” state and more weight on the “down” state. Hence, under each of these probabilities $q \in [0, q^*]$, the stock’s (normalized) current value exceeds its expected future value, i.e. $S/R^{-1} > [quS + (1 - q)dS]$. In other words, the stock has a price process which is a super martingale because it cannot be sold short.

32.6 Other Types of Frictions

Due to limitations of space, we have so far considered primarily the case of no short sales. In this section, we briefly outline the impact of other types of frictions.

Consider an alternative formulation of a short sales constraint where the admissible extent of short sales of an individual asset varies with the value of the investor’s portfolio and with any collateral pledged. Such a constraint recognizes that some assets (such as a very liquid, short-term Treasury bill) are judged to have “high” value as collateral, and thus better afford the ability to maintain a short position than is the case with other assets (such as an illiquid, off-the-run Treasury bond) deemed to have “low” collateral value. In such a setting, Hindy (1995) proved that the absence of arbitrage implies that every asset’s price admits a decomposition into a dividend-based value and a residual that depends on the asset’s “collateralizability.” Thus, the law of one price may not hold: asset k may sell at a higher price than asset l even if their payoffs are the same if a one dollar worth of asset k allows investors the ability to short more of a third asset j than does a dollar worth of asset l .⁵

Transactions costs in trading some or all assets constitute yet another type of market friction. In a binomial stock price model with proportional transactions costs, Bensaid, et al. (1992), showed that even when an option’s payoff can be exactly replicated, it can be cheaper to hedge an option with a strategy that results in a payoff that dominates that of the option when there are transactions costs. This result is foreshadowed in Boyle and Vorst (1992) who derive the cost of exactly hedging an option in an identical framework, and show that their hedge portfolio’s cost is increasing in the number of trading periods for a high enough transaction cost parameter, and for options close to at-the-money—i.e. those which have a lot of convexity and whose exact replication requires a lot of rebalancing. Thus, the

intuition from these papers is essentially that the benefits of exact replication can be traded off against savings on transactions costs. It should also be intuitively clear that in such settings that the cost of super-replicating a pool of payoffs may be cheaper than the sum of the costs of super-replicating the individual payoffs. In other words, the sublinearity result of Proposition 2 will continue to hold.

32.7 Conclusion

We have provided a review of the principal results which obtain when there are no arbitrage opportunities in a world where investors have to contend with market frictions. We conclude with some remarks about the consistency of these results with equilibrium.

One of the advantages of the no-arbitrage approach to valuation is that it allows one to make predictions about prices that are independent of particular investor attributes such as risk aversion, endowments etc. The reason is that the prices of the existing primitive assets effectively subsume the risk preferences of the marginal investor. Furthermore, in the absence of frictions, all investors' marginal utility-based valuations of all traded assets coincide: i.e. any investor may be taken to be the marginal agent supporting prices.

When there are frictions, investors' valuations may be heterogeneous, and hence differ from that predicted by the no-arbitrage approach. For instance, when there are short sales constraints, Chen (1995) showed that the price of a security derived from the no-arbitrage condition may be lower than the price that the seller of the security can actually receive by selling it to the investor who values it most. Furthermore, as Detemple and Murthy (1997) showed, the introduction of what may otherwise be considered redundant securities can upset a given equilibrium in the presence of constraints on portfolio weights. More recently, Hara (2000) shows that even when introduction of a new security does not change utility-maximizing consumption choices it may give rise to a multiplicity of each investor's security demands which in turn raises subtle equilibrium issues.

Thus, while routine application of the no-arbitrage approach in the presence of market frictions is not necessarily as useful as in a frictionless world, it nevertheless presents exciting new challenges for future research in asset pricing.

Notes

1. Some other papers relevant to arbitrage and market frictions, which we do not discuss are Dybvig and Ross (1986), Jarrow and O'Hara (1989), and Prisman (1986).
2. Other important types of market frictions include (a) a constraint on portfolio weights (such as that under a leverage constraint or margin restriction) where the permitted amount of short sales or borrowing varies with the value of the portfolio, (b) unlimited short sales at a cost that increases with the extent of

- short sales, and (c) transactions costs that have either or both a fixed component and a variable component.
3. Given the availability of a risk-free bond, every payoff has such a minimum cost. Also note that each primitive asset must satisfy $\phi(D_j) = S_j(0)$, $j = 1, \dots, J$, for if this were not true, they would not be held by any investor (which is incompatible with the fact that they are in positive net supply).
 4. In this finite dimensional setting, the new probability measure associated with $\phi^{NS}(\cdot)$ need not be equivalent to P ; i.e. the new measure need not assign positive probabilities to the same states that P does. However, limiting arguments can be used in an infinite state space to establish equivalency.
 5. Note that such a violation of the law of one price does not occur in Sects. 32.4 and 32.5 where we considered a simpler type of short sales constraint.

References

- Bensaid, B., J.P. Lesne, H. Pages, and J. Scheinkman. 1992. Derivative asset pricing with transactions costs. *Mathematical Finance* 2: 63–86.
- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–654.
- Boyle, P., and T. Vorst. 1992. Option pricing in discrete time with transactions costs. *Journal of Finance* 47: 271–293.
- Chen, Z. 1995. Financial innovation and arbitrage pricing in frictional economies. *Journal of Economic Theory* 65: 117–135.
- Cox, J., and S. Ross. 1976. The Valuation of options for alternative stochastic processes. *Journal of Financial Economics* 3: 145–166.
- Detemple, J., and S. Murthy. 1997. Equilibrium asset prices and no-arbitrage with portfolio constraints. *Review of Financial Studies* 10: 1133–1174.
- Dybvig, P., and S. Ross. 1986. Tax clienteles and asset pricing. *Journal of Finance* 41: 751–762.
- . 1987. Arbitrage. In *The New Palgrave: A dictionary of economics*, ed. J. Eatwell, M. Milgate, and P. Newman, vol. 1, 100–106. New York: Stockton Press.
- Garman, M., and J. Ohlson. 1981. Valuation of risky assets in arbitrage-free economies with transactions costs. *Journal of Financial Economics* 9: 271–280.
- Hara, C. 2000. Transaction costs and a redundant security: divergence of individual and social relevance. *Journal of Mathematical Economics* 33: 497–530.
- Harrison, M., and D. Kreps. 1979. Martingales and arbitrage in multiperiod security markets. *Journal of Economic Theory* 20: 381–408.
- Hindy, A. 1995. Viable prices in financial markets with solvency constraints. *Journal of Mathematical Economics* 24: 105–136.
- Jarrow, R., and M. O'Hara. 1989. Primes and scores: an essay on market imperfections. *Journal of Finance* 44: 1263–1287.
- Jouini, E., and H. Kallal. 1995a. Martingales and arbitrage in securities markets with transaction costs. *Journal of Economic Theory* 66: 178–197.
- . 1995b. Arbitrage in securities markets with short-sales constraints. *Mathematical Finance* 5: 197–232.
- Luttmner, E. 1996. Asset pricing in economies with frictions. *Econometrica* 64: 1439–1467.
- Merton, R. 1973. Theory of rational option pricing. *Bell Journal of Economics and Management Science* 4: 141–183.

- Modigliani, F., and M. Miller. 1958. The cost of capital, corporation finance and the theory of investment. *American Economic Review* 48: 261–297.
- Prisman, E. 1986. Valuation of risky assets in arbitrage free economies with frictions. *Journal of Finance* 41: 293–305.
- Ross, S. 1976. Return, risk and arbitrage. In *Risk and return in finance*, ed. I. Friend and J. Bicksler, 189–218. Cambridge, MA: Ballinger.
- . 1978. A simple approach to the valuation of uncertain income streams. *Journal of Business* 51: 453–475.

Fundamental Tradeoffs in the Publicly Traded Corporation

33

Joseph P. Ogden

Abstract

This chapter discusses some fundamental cost-benefit tradeoffs involving publicly traded corporations from a corporate finance viewpoint. The fundamental benefits include greater access to capital at a lower cost and economies of scale. The potential costs are associated with two fundamental problems: principal–agent conflicts of interest and information asymmetry. Various mechanisms have evolved in the United States to mitigate these problems and their costs, so that the bulk of the fundamental benefits can be realized.

Keywords

Contract devices · Conflicts of interest · Cost of capital · Disclosure · Economies of scale · Information asymmetry · Intermediaries · Liquidity · Monitoring · Signaling

33.1 Introduction

This chapter discusses, from a corporate finance perspective, the fundamental benefits and costs associated with the publicly traded corporation as a form of business organization. The fundamental benefits are two-fold. First, by incorporating and attaining public-trading status a firm gains access to a large pool of capital, which it can use to pursue capital investment projects that take advantage of economies of scale. Second, a firm's cost of capital is reduced because public investors will accept a lower cost of capital, and this is so because investors are diversified and the firm's securities are more liquid.

J. P. Ogden (✉)

State University of New York at Buffalo, Buffalo, NY, USA

e-mail: joeogden@buffalo.edu

Costs relate to two fundamental problems that beset the publicly traded corporation, both of which are consequences of the separation of ownership and control. The first problem involves “principal–agent conflicts of interest.” The second problem is “information asymmetry.” This chapter discusses these fundamental problems, their potential costs, and various mechanisms that have evolved in the United States to mitigate these problems and their costs, so that the bulk of the fundamental benefits can be realized.

33.2 Fundamental Benefits of the Publicly Traded Corporation

The fundamental benefits of the publicly traded corporation are two-fold. First, by attaining public-trading status, a firm gains access to the large pool of equity capital that is available in the public equity markets, and also enhances its access to credit markets for debt capital. Large amounts of capital allow a firm to pursue capital investment projects that take advantage of economies of scale, and thus are more profitable. Second, as many corporations emerge, secondary markets develop that allow investors to trade corporate securities and become diversified. In addition, secondary markets increase the value of corporate securities by increasing their liquidity and decreasing the cost of debt and equity capital, which in turn increase the assessed profitability of corporate projects.

33.2.1 Economies of Scale

All for-profit businesses are established to create value. The corporation is specially designed to create value on a large scale. A corporation is a separate legal entity, tethered to its owners by shares of stock. The two basic legal characteristics that distinguish a corporation from other forms of business (e.g. a sole proprietorship) are “limited liability” and the “separation of ownership and control.” Regarding the first, the extent of stockholders’ financial responsibility for the liabilities of a corporation that they collectively own is limited to the corporation’s assets, and does not extend to the stockholders’ personal assets. Regarding the second, in most corporations ownership is vested in one group, stockholders, while control is vested in another group, management (though, of course, managers may hold some of the firm’s shares).

These two legal characteristics allow a corporation to create value efficiently and on a large scale. Limited liability allows many individuals to pool their capital without concern for legal complexities and inefficiencies that would be involved if the personal assets of each individual were involved. As Jensen and Meckling (1976) explain, with unlimited liability individual stockholders would need to monitor each other’s wealth in order to estimate their own liability, which would be very costly if the firm’s shares were widely held.

The separation of ownership and control allows the two basic inputs in any economy, capital and expertise, to be contributed by separate individuals. Some

individuals have expertise to develop and undertake profitable real investments, but lack capital, whereas other individuals have capital, but lack the time and/or expertise to undertake profitable real investments. The corporation combines these two factors of production under a formal efficient structure.

Moreover, economies of scale are present in virtually all business activities, and are generally very large. Scale economies allow a larger firm, at least potentially, to create substantial value by reducing the cost of production. A corporation has the potential to amass large amounts of capital, which in turn allows it to pursue capital investment projects that take advantage of economies of scale, and thus are more profitable.

33.2.2 Reducing the Cost of Capital: Diversification and Liquidity

Two additional important benefits are associated with the publicly traded corporation: diversification and liquidity. To see these benefits, note that each firm in the economy can amass a large amount of capital by appealing to many investors to become stockholders. In turn, each investor can invest only a small portion of his or her investable wealth in any given firm, and therefore can invest in the equities of many firms simultaneously. Thus, investors can reduce the risk of their portfolios by diversifying across many firms. Risk-averse investors will accept a lower expected return on the equity of each firm because they can eliminate much of the risk of these investments via diversification. Consequently, each firm's cost of equity capital will be lower than would be the case if investors were not diversified. In turn, if all firms in the economy face a lower cost of equity capital, more projects will be deemed profitable (i.e., value-creating).

A security is liquid to the extent that an investor can quickly buy or sell the security at or near a fair price and at a low transaction cost. Liquidity is important to an investor because the ultimate purpose of investment is to provide for future consumption, either sooner or later. Investors will accept a lower expected return on equity (and thus firms will enjoy a lower cost of equity capital) if equities are liquid. The liquidity of securities naturally follows from investors' desire to become diversified. This is so because secondary markets will develop to allow trading in securities. (For additional discussion, see Ogden et al. 2003, pp. 76–77).

33.3 Fundamental Costs of the Publicly Traded Corporation

According to Modern Corporate Finance Theory, two fundamental problems beset the publicly traded corporation: “principal–agent conflicts of interest and information asymmetry.” These problems are important because they can potentially impose costs that are sufficiently large as to threaten the fundamental benefits discussed in the previous section. This section discusses these problems and their costs. The next section discusses various mechanisms designed to alleviate these problems, and thus to mitigate their costs.

33.3.1 Principal-Agent Conflicts of Interest

The first fundamental problem concerns the relationship between a principal and an agent. In general, a principal hires an agent to act in the principal's interest by performing a specified task. A problem arises in that the agent is hired to act in the principal's interest and yet, as a self-interested human being, the agent cannot be expected to subordinate his or her own interests. Thus, a conflict of interest naturally arises between the principal and the agent. In corporate finance, two types of principal-agent conflicts of interest are paramount: (1) conflicts between a firm's stockholders (as principals) and its management (as agents); and (2) conflicts between the firm (as an agent) and its creditors (as principals).

Regarding conflict (1), a firm's management is hired to act in the stockholder's interest, which is generally assumed to maximize the market value of the firm's equity. However, managers are ultimately interested in maximizing their own welfare and they control the firm. As noted earlier, an important feature of the corporation is the separation of ownership and control, and this feature is critical to capturing the stated fundamental benefits.

Managers have a self-serving incentive to capture "private benefits of control." The following are among the activities that management might employ to capture such benefits: (a) excessive consumption of "perquisites," (b) manipulating earnings and dividends, (c) maximizing the size of the firm, rather than the market value of its equity, (d) siphoning corporate assets, (e) excessive diversification at the corporate level, (f) a bias toward investments with near-term payoffs, (g) underemployment of debt, (h) entrenching their positions, and (i) packing the firm's board of directors with cronies. (For a discussion of these activities, see Ogden et al. 2003, pp. 83–88.) In the absence of mechanisms (discussed later) to offset management's private incentives, the costs to stockholders of such activities can be sufficiently large as to negate the stated fundamental benefits of the publicly traded corporation.

In addressing conflict (2), we generally assume that the conflict of interest between stockholders and management is resolved. Instead, we focus on a conflict of interest between the firm and its creditors. In this context, a creditor can be seen as a principal who "hires" the firm as an agent (i.e. by paying money up-front in the form of a loan) to act in the creditor's interest (i.e. to operate the firm in a manner that ensures that timely interest and principal payments will be made to the creditor.)

A conflict arises in that the firm's management, who will be controlling the firm, is hired to act in the stockholders' interest, rather than the creditor's interest. In the absence of mechanisms (discussed later) to protect the creditor's interest, management may engage in any of the following activities that serve to expropriate wealth from a creditor to stockholders: (a) increasing leverage, especially by subsequently issuing additional debt that has equal priority to the firm's original debt, (b) increasing the riskiness of the firm's operations (the "asset substitution" or "risk-shifting" problem), and (c) paying dividends. (For a discussion of these activities, see Ogden et al. 2003, pp. 88–93).

In addition, Myers (1977) identifies an important deadweight cost of debt called the “underinvestment problem” or the “debt overhang problem.” If a firm has default-risky debt outstanding and a profitable investment opportunity that must be financed with equity, the firm’s management might forego the investment even though it is profitable per se. This can occur if a sufficient portion of the net present value (NPV) of the project transfers to the creditors (i.e. creditors are made better off by the adoption of the project) such that the net benefit to stockholders (i.e. net of their cash contribution) is negative.

33.3.2 Information Asymmetry

The second fundamental problem is called the “information asymmetry” problem. Akerlof (1970) is generally credited with recognizing information asymmetry as a general problem in a market, though its application to corporate finance was quickly recognized. To illustrate the problem, Akerlof refers to the market for used automobiles. The crux of the problem is that the quality of a particular used make and model of automobile varies across the units for sale, and sellers know more about the quality of their unit than do potential buyers. Sellers of low-quality units have an incentive to make minimal repairs and otherwise exaggerate the quality of their unit to mimic the better-quality units in the market. As a result, in equilibrium all units will share a common price, which reflects the (true) average quality of units for sale. However, this equilibrium is unsustainable because some or all of the sellers of (truly) better-quality units will exit the market. After they exit, the true average quality of units in the market is lower, and thus so must the common price. This process will continue until only the “lemons” remain in the market. In short, the market for used automobiles can collapse under the weight of information asymmetry (see Ogden et al. 2003, pp. 101–102).

The markets for corporate securities are also naturally beset by information asymmetry because of product market competition and the separation of ownership and control. A firm’s management (assumed to be acting in the stockholders’ interest) must devise strategic plans to compete in its chosen product market. These plans cannot be divulged to the firm’s diffuse stockholders for the simple reason that this would be tantamount to revealing the information to the firm’s competitors, who could surreptitiously become stockholders in order to gain access to such plans, and then thwart them with counter-strategies.

Consequently, a firm’s management generally and necessarily has more information about the firm’s operations, and thus its true value, than do the firm’s actual or potential stockholders. In the absence of mechanisms (discussed below) to mitigate this information asymmetry problem, the market for corporate securities may be very poor, and could even collapse à la Akerlof’s argument.

33.4 Mitigating the Costs

A wide variety of mechanisms have emerged in the U.S. and other markets to mitigate costs associated with both the agency and information asymmetry problems discussed above. This section briefly discusses such mechanisms, following a top-down approach. (For an in-depth discussion of each of these mechanisms, see Ogden et al. 2003, Chap. 5).

33.4.1 Government Laws and Regulations

The most fundamental services that government provides are for establishment of property rights through legislation and the enforcement of legal contracts through a judicial system. In addition, government generally regulates the financial markets, such as with the US Securities Act of 1933 and the Securities Exchange Act of 1934, the latter of which established the Securities and Exchange Commission (SEC). The SEC requires all firms with publicly traded securities to register such securities, to file periodic financial statements, etc. The 1934 act also requires publicly traded firms to provide stockholders the opportunity to vote on matters such as the election of board directors. The SEC also prohibits insider trading, requires major owners of a firm's equity to disclose their ownership, etc. The regulations imposed by the SEC most obviously help to reduce information asymmetry and associated costs. In addition, these regulations curb the self-serving activities of managers, and thus help to mitigate costs associated with principal-agent conflicts of interest.

33.4.2 Securities Traders, Analysts, and the Press

Securities traders, analysts, and the press all generate important information about the values of securities and the efficacy of firms' managements. Their efforts help to reduce information asymmetry. In addition, they serve as indirect monitors of each firm's management, which curbs management's opportunities to engage in self-serving activities, and thus mitigates costs associated with stockholder-management conflicts of interests.

33.4.3 Ownership Structure

A firm's stockholders can promote their own interest through "activism," specifically, by voting on major management-initiated proposals, submitting their own proposals, and monitoring management's decisions. Unfortunately, if a firm's ownership is diffuse, activism is generally thwarted by the "free-rider problem," whereby

stockholders have an incentive to freely benefit from the costly efforts of others to monitor and reform management.

One means of mitigating the free-rider problem is for a firm's equity to be closely held (i.e. by nonmanagement stockholders), so that most or all of a firm's stockholders have a sufficient financial incentive to monitor management. However, close ownership is costly because it reduces the benefits of the corporate form discussed earlier. Alternatively, a firm may require management to own a substantial number of the firm's shares (or to hold stock options), which serves to align stockholders' and management's interests. However, this mechanism is also imperfect. For instance, if the bulk of a manager's personal portfolio is invested in the manager's own firm, his or her financial policies (i.e. real investment and financial decisions) may be distorted in a way that reduces the value of the firm's shares.

33.4.4 Board Oversight

Board directors are hired to protect and promote the interests of a firm's stockholders. The existence of boards of directors is perhaps the most obvious indication of a potential conflict of interest between stockholders and management. An independent board can be an effective advocate of stockholders' interests because the board generally has powers to: (1) require board approval of major capital expenditures, acquisitions, divestitures, and security offerings, (2) control the firm's capital structure, (3) hire outside consultants to scrutinize major projects, and (4) as necessary, fire senior management.

However, senior management has an incentive to "pack the board" with its cronies. If management is successful in doing so, the board becomes little more than a "rubber stamp" for management's decisions. To avoid this, stockholders should insist on an independent board, consisting of mostly outsiders who are not beholden to management.

33.4.5 Financial Institutions

A wide variety of financial institutions exist in the United States, including commercial banks, finance companies, insurance companies, venture capital firms, and securities firms (i.e. underwriters). According to theory, one of the most fundamental services provided by financial institutions is to mitigate costs associated with the information asymmetry problem in the market for corporate securities. Sellers of valuable proprietary information work through a financial institution that acts as an "intermediary" between the firm and investors. The intermediary is in a position to verify the value of the proprietary information, and yet can be trusted to keep such information confidential.

33.4.6 Contract Devices

Finally, firms employ a variety of contract devices to mitigate principal–agent conflicts of interest and/or information asymmetry problems. To illustrate, we briefly discuss two types of contracts that can alleviate conflicts: executive compensation contracts and debt contracts.

For a firm with diffuse ownership, devising a contract with senior management is problematic because managers have private incentives to maximize their own welfare, as discussed earlier. In order to align the interests of stockholders and management, a firm may include performance-based provisions in the executive's compensation contract, such as an annual earnings-based bonus or grants of stock or stock options. Such provisions may serve the intended purpose reasonably well, but could also backfire by causing management to distort the firm's capital investment program, its capital structure, its dividend policy, etc.

We also mentioned earlier that a conflict of interest arises between a borrowing firm and its creditors, whereby the firm has an incentive to take actions to expropriate wealth from creditors. Creditors can mitigate these incentives by including various provisions and covenants in the debt contract. For instance, a creditor may demand collateral to mitigate the asset substitution problem. Creditors may also restrict the firm's ability to issue additional debt and to restrict or prohibit the payment of dividends.

33.4.7 Signaling

Finance theory also suggests that firms can provide “signals” of true value to the market in order to mitigate the information asymmetry problem. In the finance literature, authors have suggested each of the following signaling mechanisms: ownership structure, leverage, dividends, and stock repurchases. (For discussion, see Ogden et al. [2003](#), Chap. 4).

33.5 Summary

This chapter discusses fundamental tradeoffs associated with the publicly traded corporation. On the positive side, corporations allow for the concentration of large quantities of capital, which can be used for investing in large capital investment projects that capture valuable economies of scale. In addition, the corporate form allows investors to diversify and to trade securities, which reduces the cost of capital. On the negative side, the separation of ownership and control engenders two fundamental and potentially costly problems: principal–agent conflicts of interest and information asymmetry. Various mechanisms have evolved at various levels to mitigate these problems and their costs, so that more of the fundamental benefits of the publicly traded corporation can be realized.

References

- Akerlof, G. 1970. The market for 'lemons,' qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics* 84: 488–500.
- Jensen, M.C., and W. Meckling. 1976. Theory of the firm: managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics* 1: 305–360.
- Myers, S.C. 1977. Determinants of corporate borrowing. *Journal of Financial Economics* 5: 146–175.
- Ogden, J.P., F.C. Jen, and P.F. O'Connor. 2003. *Advanced corporate finance: Policies and strategies*. Upper Saddle River: Prentice-Hall.

The Mexican Peso Crisis

34

Fai-Nan Perng

Abstract

The Mexican Peso Crisis was the byproduct of various developments including large inflows of short-term foreign capital, prolonged current account deficit, and political instability. Between 1990 and 1993, investors in the United States were particularly eager to provide loans, many of them short-term, to the Mexican government and to Mexican corporations. Throughout this period, the share of foreign capital inflows exceeded the current account deficit. However, political instability and U.S. interest rate hikes soon changed the optimism for Mexico's economic outlook. At the beginning of 1994, this did not affect the value of the peso, for Mexico was operating with a target zone exchange rate and its central bank stood ready to accept pesos and pay out dollars at the fixed rate. Yet Mexico's reserves of foreign currency were too small to maintain its target zone exchange rate. When Mexico ran out of dollars at the end of 1994, the Mexican government announced a devaluation of the peso. As a result, investors avoided buying Mexican assets, adding to downward pressure on the peso.

Overall, the Mexican meltdown of 1994–1995 had many facets. Yet couple of lessons are particularly clear: while foreign capital can make up for the shortfall in domestic saving, only long-term capital – in the forms of foreign direct investment or long-term debt – is conducive to domestic investment; large and abrupt movements of capital across national borders can cause excessive financial market volatility and undermine economic stability in the countries involved. Last and most importantly, prolonged current account deficit should be remedied by allowing the domestic currency to depreciate, promoting savings, or cutting back government expenditure rather than financed by foreign capital inflows.

F.-N. Perng (✉)

Central Bank of the Republic of China (Taiwan), Taipei, Taiwan, Republic of China

Countries with protracted current account deficits such as Argentina, the Philippines, Indonesia, Thailand, and Saudi Arabia, with Thailand in particular, should heed Mexico's experience.

Keywords

Current account deficit · Foreign capital inflow · Foreign exchange reserves · Long-term debt · Short-term debt · Target zone exchange rate

The Mexican government stopped repaying external debt obligations in August 1982 due to a shortfall of its foreign exchange reserves. Brazil, Argentina, and Chile followed suit, which triggered what came to be known as the Latin American Debt Crisis. Later, following a series of economic and financial reforms, conditions in Mexico gradually improved sufficiently to start attracting inward investment again. The debt relief initiative put forth by U.S. Treasury Secretary Nicholas Brady in 1990 eventually put Mexico more firmly back on her feet. Thereafter, Mexico engaged in several rounds of negotiations with the United States with the intent of securing the North American Free Trade Agreement and eventually won the U.S. Congress over toward the end of 1993. It is fair to say that in the years between 1990 and 1993, most foreign investors were bullish about the outlook of the Mexican economy.

In a separate development, the advancement of telecommunications and computer technology has sped and immensely reduced the cost of transferring capital. Moreover, a growing number of households started to entrust their savings with professional fund managers. Portfolios managed by fund managers tend to be well diversified with assets invested in multiple currencies (huge sums of money can literally be moved from one currency to another or one financial product to the next at the push of a button), a practice that further hastened the speed and amplified the magnitude of international capital movements (the combined value of cross-border portfolio investment in Europe, America, and Japan reached \$2,500 billion in 1991). While this was taking place, the U.S. economy was mired in a protracted downturn. The Federal Reserve rightly countered with a monetary stimulus. The interest rate on the 3-month fixed deposits was slashed from 10.25% in March 1989 to 3.1875% in January 1994. Against the backdrop of highly efficient international financial markets, U.S. investors moved a huge chunk of their capital abroad. A significant portion of this outflow was absorbed by emerging markets in Latin America including, of course, Mexico.

Owing to the various developments outlined in the preceding paragraphs, substantial foreign capital began to flow into Mexico in 1990. The size of foreign capital inflow ballooned from \$8.441 billion in 1990 to \$32.06 billion by 1993. Altogether, some \$92.647 billion of foreign capital swamped Mexico's financial markets in those 4 years (Table 1).

Among the first to be affected by this foreign capital inflow was, not surprisingly, the Peso. At that time, Mexico's exchange rate system was one of target zone. The

Table 1 Mexico's balance of payments (1990–1994)

Items	1990	1991	1992	1993	1994 ^a	
Current account	-7,451	-14,888	-24,806	-23,391	-7,020 (Jan.–Nov.)	
Trade balance	-4,433	11,329	20,677	18,891		
Services	-6,993	-6,305	-7,150	-7,187		
Transfers	3,975	2,746	3,021	2,687		
Capital account	8,441	25,139	27,008	32,059	14,600 (Jan.–Nov.)	
Direct investment	2,549	4,742	4,393	4,901		
Portfolio investment	-3,985	12,138	19,175	27,867		
Other investment	9,877	8,259	3,440	-709		
Government borrowings	1,657	-1,454	-5,867	-1,136		
Net errors and omissions	1,228	-2,278	-5,867	-1,436		
Reserves and related items ^b change in reserve assets	-2,218	-7,973	-1,745	-7,232		
(-: increase)	-3,479	-7,834	-1,118	-6,129		
Foreign exchange reserves (year end)	9,446	17,140	18,394	24,886	16 Dec. 94	11,150
					13 Jan. 95	3,480
Exchange rate (year end, Peso/US\$)	2,9454	3,0710	3,1154	3,1059	19 Dec. 94	22 Dec. 94
					31 Jan. 95	6 Feb. 95
						5,3350

Source: International financial statistics, published by IMF on Jan. 1995

^aA *plus sign* indicates a reduction in assets or an increase in liabilities; a *minus sign* indicates the opposite^bAs published by Mexico's central bank in its monetary policy report on 25 January 1995

lower bound of the Peso/USD exchange rate had been set at 3.05 since November 1991 when foreign exchange controls were removed. The upper bound had been raised gradually at a rate of 0.0004 Peso per day, beginning on 21 October 1992. The idea was to allow the Peso/USD exchange rate to adjust within a band wide enough to properly reflect supply and demand in the foreign exchange market (Fig. 1).

Before the end of 1993, the Peso/USD exchange rate was relatively stable due to large and sustained foreign capital inflows that more than offset current account deficits. Under this arrangement, the integrity of Mexico's target zone exchange rate regime was not put to test. At the same time, the Mexican inflation rate was running at a significantly higher level than that of the United States. Between 1990 and September 1994, the U.S. CPI rose by only 4%. During the same period, the Mexican CPI jumped by 61.3%. In a parallel development, the Peso depreciated from 2.9454 to 3.4040 to a dollar. According to the purchasing power parity, the Peso/USD exchange rate should have been 4.5682 in September 1994. In other words, the Peso was overvalued by 34% (Table 2 and Fig. 2).

Maintaining a stable Peso/USD exchange rate helped to push the Mexican inflation down, as American prices were stable. Mexican CPI inflation was 23.3% in 1990, which dropped to 8.4% by September 1994. An overvalued Peso, however, undermined the competitiveness of Mexican exports. It's a small wonder that the position of the current account continued to worsen. A deficit of \$7.451 billion in 1990 swelled to \$23.391 billion in 1993, a figure approaching 6% of Mexico's GDP. For 1994, this figure was projected to rise to \$28 billion or 8% of Mexico's GDP (Table 1).

Throughout this period, the share of foreign capital inflows that exceeded the current account deficit was bought by the central bank. This would explain why Mexico's foreign exchange reserves rose from \$5.946 billion in 1989 to \$24.886 billion by the end of 1993 (Table 1).

At the beginning of 1994, for a variety of reasons, investor confidence began to wane. External factors include U.S. interest rate hikes that started from February 1994 and the resulting rise in the rate of return from investing in the dollar, which were inarguably the most important. Explanations that had roots at home include the January 1994 riot in the southern province of Chiapas, the assassination of Señor Colosios, the ruling party's presidential candidate in March, and the deterioration of the current account deficit. No longer upbeat about the prospect of the Mexican economy and recognizing that the Peso/USD exchange rate had become unstable, foreign capital inflow dried up to a level that could no longer sustain the current account deficit. Demand for the U.S. dollar far exceeded the supply in the foreign exchange market; the Peso sapped. In order to keep the Peso exchange rate within the upper bound of the target zone, the Mexican government intervened heavily by selling the dollar, a process that quickly depleted precious foreign exchange reserves. In order to replenish official reserves that were running at a dangerously low level, Mexico issued the Tesobonos, a U.S. dollar-denominated short-term debt payable in Peso at maturity. In retrospect, Mexico's central bank should have tightened its monetary stance. But instead, it acted like an innocent bystander, fearing such measures would dampen economic growth and hoping that foreign investors will

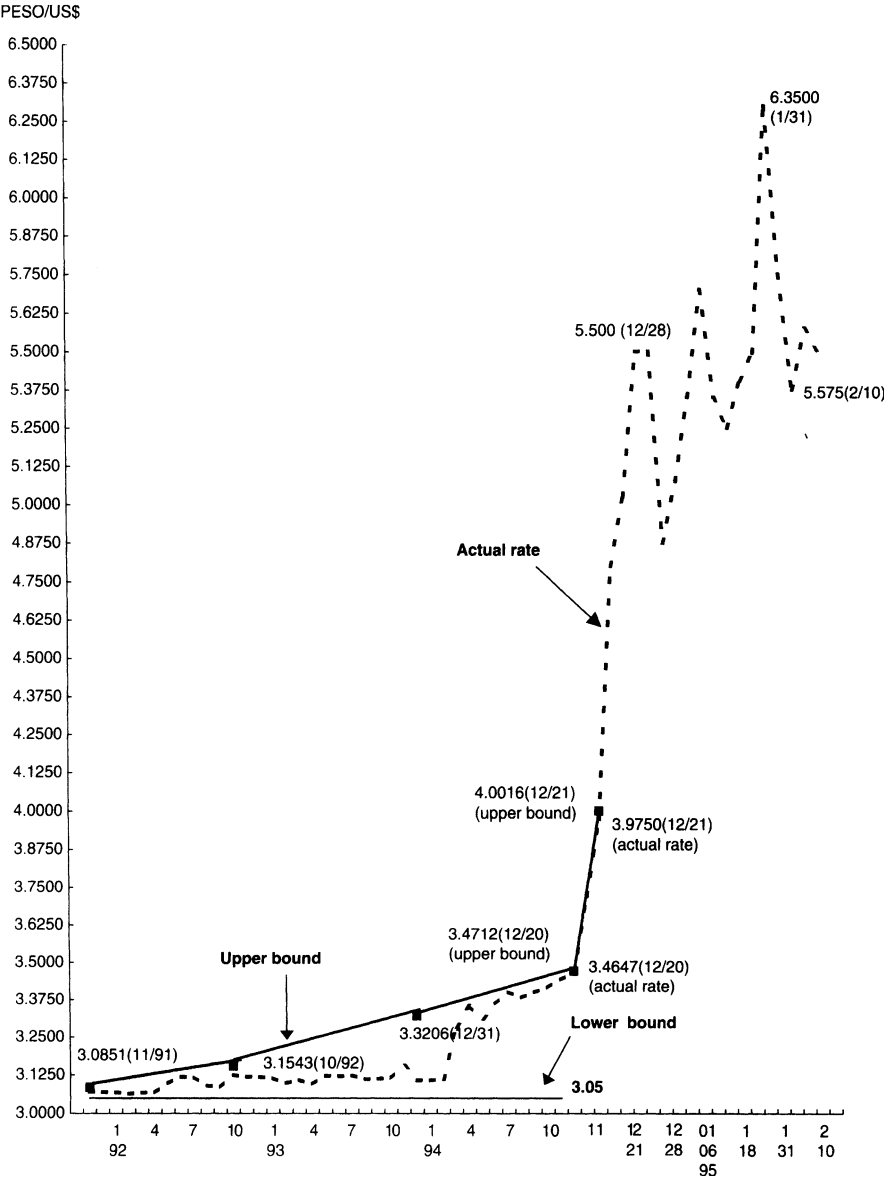


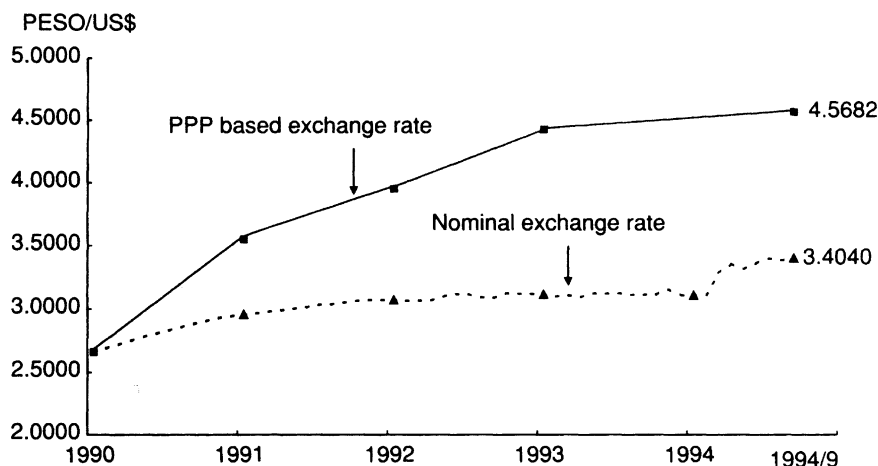
Fig. 1 Peso/US\$ exchange rate

return in droves after the presidential election in August. Although the ruling party did get re-elected, its secretary general, Señor Masssieu, was assassinated in September. When the newly elected President Zedillo was sworn into office on December 1, Mexico's financial markets were on the brink of collapse.

Table 2 Peso/US\$ exchange rate and inflation comparison

	1990	1991	1992	1993	1994/9
Nominal exchange rate (Peso/US\$)	2.9454	3.0710	3.1154	3.1059	3.4040
PPP exchange rate (Peso/US\$)	2.9454	3.5421	3.9944	4.2842	4.5682
Whole Sale Price Index (WPI)	100.0	120.5	136.7	148.8	163.3
Mexico	(23.3)	(20.5)	(13.4)	(8.9)	(8.4)
Producer Price Index (PPI) US	100.0	100.2	100.8	102.3	104.0

Note: annual growth rate % in brackets

**Fig. 2** Nominal exchange rate and purchasing power parity (PPP) based exchange rate

Among the many forces that were weighing on Mexico's financial markets, the overvalued Peso and the accompanying deterioration in the current account deficit were the most obvious. Another key feature of the Mexican debacle was that foreign capital inflows had predominantly been invested in short-term debts. Of the \$32 billion worth of capital inflow in 1993, \$27.9 billion was invested in such instruments (Table 1). The source of this type of foreign capital inflow can be traced to mutual funds (Fidelity alone had \$8 billion invested in emerging markets in 1994), and hedge and pension funds. All it took was one telephone call for the funds to be shifted out of Mexico once the fund managers convinced themselves that the Peso exchange rate was unstable or when the sentiment on the outlook of the Mexican economy suddenly turned bearish.

With the current account position worsening and the inflow of foreign capital reduced to a trickle, the Mexican government resorted to financing the current account deficit with official reserves in addition to issuing short-term dollar debts and the Tesobonos. Mexico's foreign exchange reserves declined from \$24.886 billion at the end of 1993 to \$11.15 billion on 16 December 1994. Although Señor Serra, the finance minister, repeatedly reassured foreign investors that the upper bound of the peso exchange rate would not be breached, that very ceiling was hastily

raised from 3.4712 to 4.0016 on 20 December. The Peso exchange rate fell sharply from 3.4647 at the close of the business day on December 20 to 3.9750 the next day, getting uncomfortably close to the 4.0016 mark. Unable to stem the tide of foreign capital outflow and with the level of foreign exchange reserves running precariously low, the Mexican government had little choice but to let the Peso float (Fig. 1). The Peso fell to an all-time low of 5.5 to a dollar on 28 December.

The exodus of foreign capital not only exerted a severe downward pressure on the Peso exchange rate but also depressed stock prices. The Mexican stock index fell from 2857.5 on 23 September 1994 to 1935.32 on 9 February 1995, or 32% in 4 months (Fig. 3).

For a variety of reasons, the United States came to Mexico's rescue and brought the international financial community to the negotiating table. Possible explanations for the action taken by the U.S. government include:

1. As much as \$53 billion of debt was about to become due at the end of 1995. Unaided, Mexico was in eminent danger of repeating the 1982 crisis.
2. Mexico had become United States' third largest trading partner, with bilateral trade amounting to \$100 billion per annum. A further deterioration in the Mexican economy was more than likely to have a negative impact on the United States; the number of illegal immigrants waiting to cross the border could rise considerably.
3. The contagion effect of the Mexican crisis was beginning to be felt by other large Latin American debtors such as Brazil and Argentina. Helping Mexico would prevent the contagion from spreading further afield.
4. The aid package included broad based economic stabilization measures (putting a 7% cap on wage increases, cutting back government expenditure, and curbing the expansion of bank credits and money supply).

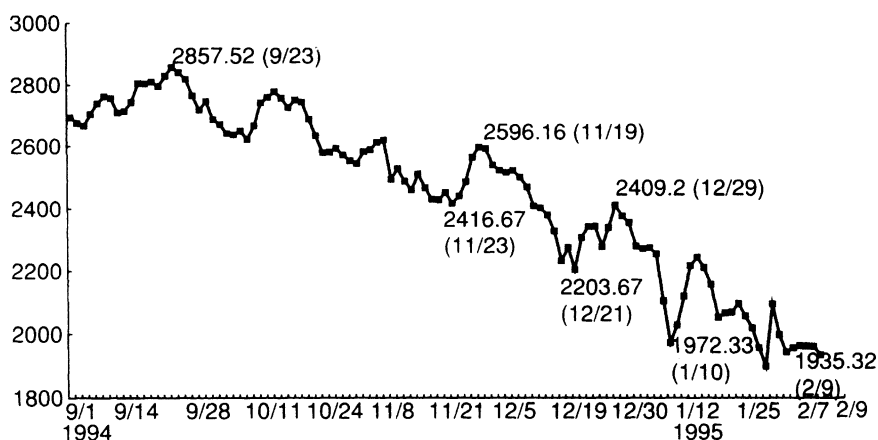


Fig. 3 Mexican stock index

President Bill Clinton's proposal to provide a \$40 billion loan guarantee that would have enabled Mexico to raise fresh capital in international financial markets and resume debt repayment was rejected by the U.S. Congress on 30 January. The Peso took the hit and tumbled to 6.35 to a dollar the next day. By then, Mexico had only enough foreign exchange reserves to last two more days. In an emergency session, the United States, Germany, and France finally agreed to provide Mexico with a \$48.8 billion refinancing package, the details of which are as follows:

1. The U.S. government would establish a \$20 billion credit line (with \$1.4 billion coming from the Exchange Stabilization Fund of the Department of Treasury) made up of:
 - (a) A Peso/dollar swap line for maturities that fall within 12 months or between 3 and 5 years
 - (b) Guarantee for debts with maturities up to 10 years designed to help Mexico to raise new debts in international markets.

The remaining \$600 million came in the form of a temporary short-term swap credit line set up by the U.S. Federal Reserve.

2. The IMF offered a \$17.8 credit line of which \$7.8 billion came in the form of emergency credits and a \$10 billion stand-by credit facility financed by emerging market economies with ample foreign exchange reserves.
3. The Bank for International Settlements chipped in with a \$10 billion credit line consisting of swap facilities offered by its 29 member central banks (including the United States, Japan, Germany, UK, and France).
4. Argentina, Mexico, Chile, and Colombia collectively offered a \$1 billion credit line.

The combined value of the four credit lines listed above summed to \$48.8 billion. Meanwhile, a consortium of private sector financial institutions headed by Citibank and JP Morgan negotiated for a syndication loan worth \$3 billion.

The following lessons can be learnt from the Mexican financial crisis:

1. Capital formation can promote economic growth, but the most reliable source of fund is domestic savings. While foreign capital can make up for the shortfall in domestic savings, only long-term capital, in the forms of foreign direct investment or long-term debt, is conducive to domestic investment. Foreign portfolio investment channels funds into secondary markets, resulting in the transfer of ownership, but brings little direct benefit to domestic capital formation.
2. The size of global portfolio investment has grown exponentially. Fund managers make investment decisions based on predictions about future exchange rates, interest differentials, and stock prices. Large and abrupt movements of capital across national borders can cause excessive financial market volatility and undermine economic stability in the countries involved. These adverse effects would be especially acute in small but highly open economies. For this reason, capital account liberalization should follow a gradual and orderly approach.

3. Prolonged current account deficit should be remedied (by allowing the domestic currency to depreciate, promoting saving, or cutting back government expenditure) rather than financed by foreign capital inflows. A country cannot rely on external financing indefinitely. Interestingly enough, countries like Argentina, the Philippines, Indonesia, Thailand, and Saudi Arabia have all been running current account deficits since 1987. It is worth pointing out that Thailand, in particular, relies almost exclusively on short-term capital inflows to finance her current account deficit.
4. The IMF should acquire in-depth knowledge of member economies, work with them to establish an early warning system, and make policy recommendations that would prevent the outbreak of future crises.

Mr. Perng, the Governor of Taiwan's central bank, noted the long running current account deficits for a number of Asian countries in his article published in the *Commercial Times* on 23 February 1995. He stated that the situation in Thailand was especially worrisome as its current account deficit was mainly financed by short-term financial capital inflows. Fifteen months after the publication of this chapter, the Asian financial crisis broke out with Thailand at the front line of the crisis. Fai-nan Perng (*Commercial Times*, 23 February 1995)

Methods for Portfolio Performance Evaluation

35

Lalith P. Samarakoon and Tanweer Hasan

Abstract

The portfolio performance evaluation involves the determination of how a managed portfolio has performed relative to some comparison benchmark. Performance evaluation methods generally fall into two categories, namely conventional and risk-adjusted methods. The most widely used conventional methods include benchmark comparison and style comparison. The risk-adjusted methods adjust returns in order to take account of differences in risk levels between the managed portfolio and the benchmark portfolio. The major such methods are the Sharpe ratio, Treynor ratio, Jensen's alpha, Modigliani and Modigliani, and Treynor Squared. The risk-adjusted methods are preferred to the conventional methods.

Keywords

Performance · Evaluation · Standard deviation · Systematic risk · Conventional methods · Benchmark comparison · Style comparison · Risk-adjusted measures · Sharpe measure · Treynor measure · Jensen measure · Alpha · Modigliani-Modigliani measure · Treynor squared

L. P. Samarakoon (✉)
University of St. Thomas, St. Paul, MN, USA
e-mail: lpssamarakoon@stthomas.edu

T. Hasan
Independent University of Bangladesh, Dhaka, Bangladesh
e-mail: t.hasan@iub.edu.bd

35.1 Introduction

The portfolio performance evaluation primarily refers to the determination of how a particular investment portfolio has performed relative to some comparison benchmark. The evaluation can indicate the extent to which the portfolio has outperformed or underperformed, or whether it has performed at par with the benchmark.

The evaluation of portfolio performance is important for several reasons. First, the investor, whose funds have been invested in the portfolio, needs to know the relative performance of the portfolio. The performance review must generate and provide information that will help the investor to assess any need for rebalancing of his investments. Second, the management needs this information to evaluate the performance of the manager of the portfolio and to determine the manager's compensation, if that is tied to the portfolio performance. The performance evaluation methods generally fall into two categories, namely conventional and risk-adjusted methods.

35.2 Conventional Methods

35.2.1 Benchmark Comparison

The most straightforward conventional method involves comparison of the performance of an investment portfolio against a broader market index. The most widely used market index in the US is the S&P 500 index, which measures the price movements of 500 US stocks compiled by the Standard and Poor's Corporation. If the return on the portfolio exceeds that of the benchmark index, measured during identical time periods, then the portfolio is said to have beaten the benchmark index. While this type of comparison with a passive index is very common in the investment world, this creates a particular problem. The level of risk of the investment portfolio may not be the same as that of the benchmark index portfolio. Higher risk should lead to commensurately higher returns, in the long term. This means if the investment portfolio has performed better than the benchmark portfolio it may be due to the investment portfolio being riskier than the benchmark portfolio. Therefore, a simple comparison of the return on an investment portfolio with that of a benchmark portfolio may not produce valid results.

35.2.2 Style Comparison

A second conventional method of performance evaluation called "style-comparison" involves comparison of return of a portfolio with that having a similar investment style. While there are many investment styles, one commonly used approach classifies investment styles as value versus growth. The "value style" portfolios invest in companies that are considered undervalued on the basis of yardsticks such as price-to-earnings and price-to-book value multiples. The "growth style" portfolios invest in companies whose revenue and earnings are expected to grow faster than those of the average company. In order to evaluate the performance of a value-oriented portfolio,

one would compare the return on such a portfolio with that of a benchmark portfolio that has a value style. Similarly, a growth-style portfolio is compared with a growth-style benchmark index. This method also suffers from the fact that while the style of the two portfolios that are compared may look similar, the risks of the two portfolios may be different. Also, the benchmarks chosen may not be truly comparable in terms of the style since there can be many important ways in which two similar style-oriented funds vary. Reilly and Norton (2003) provide an excellent disposition of the use of benchmark portfolios and portfolio style and the issues associated with their selection. Sharpe (1992) and Christopherson (1995) develop methods for determining style.

35.3 Risk-Adjusted Methods

The risk-adjusted methods make adjustments to returns in order to take account of the differences in risk levels between the managed portfolio and the benchmark portfolio. While there are many such methods, the most notables are the Sharpe ratio (S), Treynor ratio (T), Jensen's alpha (α), Modigliani and Modigliani (M^2), and Treynor Squared (T^2). These measures along with their applications are discussed below.

35.3.1 Sharpe Ratio

The Sharpe ratio (Sharpe 1966) computes the risk premium of the investment portfolio per unit of total risk of the portfolio. The risk premium, also known as excess return, is the return of the portfolio less the risk-free rate of interest as measured by the yield of a Treasury security. The total risk is the standard deviation of returns of the portfolio. The numerator captures the reward for investing in a risky portfolio of assets in excess of the risk-free rate of interest, while the denominator is the variability of returns of the portfolio. In this sense, the Sharpe measure is also called the “reward-to-variability” ratio. Equation (1) gives the Sharpe ratio.

$$S = \frac{r_p - r_f}{\sigma_p} \quad (1)$$

S = Sharpe ratio

r_p = return of the portfolio

r_f = risk-free rate

σ_p = standard deviation of returns of the portfolio

The Sharpe ratio for an investment portfolio can be compared with the same for a benchmark portfolio such as the overall market portfolio. Suppose that a managed portfolio earned a return of 20% over some time period with a standard deviation of 32%. Also assume that during the same period, the Treasury bill rate was 4%, and the overall stock market earned a return of 13% with a standard deviation of 20%. The managed portfolio's risk premium is $(20\% - 4\%) = 16\%$, while its Sharpe ratio, S ,

is equal to $16\%/32\% = 0.50$. The market portfolio's excess return is $(13\% - 4\%) = 9\%$, while its S equals $9\%/20\% = 0.45$. Accordingly, for each unit of standard deviation, the managed portfolio earned a risk premium of 0.50%, which is greater than that of the market portfolio of 0.45%, suggesting that the managed portfolio outperformed the market after adjusting for total risk.

35.3.2 Treynor Ratio

The Treynor ratio (Treynor 1965) computes the risk premium per unit of systematic risk. The risk premium is defined as in the Sharpe measure. The difference in this method is that it uses the systematic risk of the portfolio as the risk parameter. The systematic risk is that part of the total risk of an asset which cannot be eliminated through diversification. It is measured by the parameter known as “beta” that represents the slope of the regression of the returns of the managed portfolio on the returns to the market portfolio. The Treynor ratio is given by Eq. (2).

$$T = \frac{r_p - r_f}{\beta_p} \quad (2)$$

T = Treynor ratio

r_p = return of the portfolio

r_f = risk-free rate

β_p = beta of the portfolio

Suppose that the beta of the managed portfolio in the previous example is 1.5. By definition, the beta of the market portfolio is equal to 1.0. This means the managed portfolio has one-and-half times more systematic risk than the market portfolio. We would expect the managed portfolio to earn more than the market because of its higher risk. In fact in the example, the portfolio earned an excess return of 16% whereas the market earned only 9%. These two numbers alone do not tell anything about the relative performance of the portfolio, since the portfolio and the market have different levels of market risk. In this instance, the Treynor ratio for the managed portfolio equals $(20\% - 4\%)/1.5 = 10.67$, while that for the market equals $(13\% - 4\%)/1.00 = 9.00$. Thus, after adjusting for systematic risk, the managed portfolio earned an excess return of 10.67% for each unit of beta while the market portfolio earned an excess return of 9.00% for each unit of beta. Thus, the managed portfolio outperformed the market portfolio after adjusting for systematic risk.

35.3.3 Jensen's Alpha

Jensen's alpha (Jensen 1968) is based on the Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965), and Mossin (1966). The alpha represents the amount by which the average return of the portfolio deviates from the expected

return given by the Capital Asset Pricing Model. The CAPM specifies the expected return in terms of the risk-free rate, systematic risk, and the market risk premium. The alpha can be greater than, less than, or equal to 0. An alpha greater than 0 suggests that the portfolio earned a rate of return in excess of the expected return of the portfolio. Jensen's alpha is given by Eq. (3).

$$\alpha = r_p - [r_f + \beta_p(r_m - r_f)] \quad (3)$$

α = Jensen's alpha

r_p = return of the portfolio

r_m = return of the market portfolio

r_f = risk-free rate

β_p = beta of the portfolio

Using the same set of numbers from the previous example, the alpha of the managed portfolio and the market portfolio can be computed as follows. The expected return of the managed portfolio is $4\% + 1.5(13\% - 4\%) = 17.5\%$. Therefore, the alpha of the managed portfolio is equal to the actual return less the expected return, which is $20\% - 17.5\% = 2.5\%$. Since we are measuring the expected return as a function of the beta and the market risk premium, the alpha for the market is always 0. Thus, the managed portfolio has earned a 2.5% return above that must be earned given its market risk. In short, the portfolio has a positive alpha suggesting superior performance.

When the portfolio is well diversified all three methods, Sharpe, Treynor, and Jensen will give the same ranking of performance. In the example, the managed portfolio outperformed the market on the basis of all three ratios. When the portfolio is not well diversified or when it represents the total wealth of the investor, the appropriate measure of risk is the standard deviation of returns of the portfolio, and hence the Sharpe ratio is the most suitable. When the portfolio is well diversified, however, a part of the total risk has been diversified away and the systematic risk is the most appropriate risk metric. Both Treynor ratio and Jensen's alpha can be used to assess the performance of well-diversified portfolios of securities. These two ratios are also appropriate when the portfolio represents a sub-portfolio or only a part of the client's portfolio. Chen and Lee (1981, 1986) examine the statistical distribution of Sharp, Treynor, and Jensen measures and show that the empirical relationship between these measures and their risk proxies is dependent on the sample size, the investment horizon and market conditions. Cumby and Glen (1990), Grinblatt and Titman (1994), Kallaberg et al. (2000), and Sharpe (1998) provide evidence of the application of performance evaluation techniques.

35.3.4 Modigliani-Modigliani Measure

The Sharpe ratio is not easy to interpret. In the example, the Sharpe ratio for the managed portfolio is 0.50, while that for the market is 0.45, and we concluded that

the managed portfolio outperformed the market. The difficulty, however, is that the differential performance of 0.05 is not an excess return. Modigliani and Modigliani (1997) measure, which is referred to as M^2 , provides a risk-adjusted measure of performance that has an economically meaningful interpretation. The M^2 is given by Eq. (4).

$$M^2 = r_{p*} - r_m \quad (4)$$

M^2 = Modigliani-Modigliani measure

r_{p*} = return on the adjusted portfolio

r_m = return on the market portfolio

The adjusted portfolio is the managed portfolio adjusted such that it has the same total risk as the market portfolio. The adjusted portfolio is constructed as a combination of the managed portfolio and risk-free asset, where weights are specified as in Eqs. (5) and (6).

$$w_{rp} = \frac{\sigma_m}{\sigma_p} \quad (5)$$

$$w_{rf} = 1 - w_{rp} \quad (6)$$

w_{rp} represents the weight given to the managed portfolio, which is equal to the standard deviation of the market portfolio (σ_m) divided by the standard deviation of the managed portfolio (σ_p). w_{rf} is the weight on the risk-free asset and is equal to one minus the weight on the managed portfolio. The risk of the adjusted portfolio (σ_{p*}) is the weight on the managed portfolio times the standard deviation of the managed portfolio as given in Eq. (7). By construction, this will be equal to the risk of the market portfolio.

$$\sigma_{p*} = w_{rp} \times \sigma_p = \frac{\sigma_m}{\sigma_p} \times \sigma_p = \sigma_m \quad (7)$$

The return of the adjusted portfolio (r_{p*}) is computed as the weighted average of the returns of the managed portfolio and the risk-free rate (Eq. 8), where the weights are as in Eqs. (5) and (6) above.

$$r_{p*} = w_{rf} \times r_f + w_{rp} \times r_p \quad (8)$$

The return on the adjusted portfolio can be readily compared with the return on the market portfolio since both have the same degree of risk. The differential return, M^2 , indicates the excess return of the managed portfolio in comparison to the benchmark portfolio after adjusting for differences in the total risk. Thus, M^2 is more meaningful than the Sharpe ratio.

In the example, the standard deviation of the managed portfolio is 32% and the standard deviation of the market portfolio is 20%. Hence, the $w_{rp} = 20/32 = 0.625$

and $w_{rf} = 1 - 0.625 = 0.375$. The adjusted portfolio would be 62.5% invested in the managed portfolio and 37.5% invested in Treasury bills. Now, the risk of the adjusted portfolio, $\sigma_{p^*} = 0.625 \times 32\% = 20\%$, is the same as the risk of the market portfolio. The return on the adjusted portfolio would be $r_{p^*} = 0.375 \times 4\% + 0.625 \times 20\% = 14\%$. The $M^2 = 14\% - 13\% = 1\%$. Thus, the on a risk-adjusted basis the managed portfolio has performed better than the benchmark by 1%.

35.3.5 Treynor Squared

Another performance measure, called T^2 analogous to M^2 can be constructed. This is a deviant of the Treynor measure, and the rationale is the same as that of M^2 . T^2 is defined as in Eq. (9).

$$T^2 = r_{p^*} - r_m \quad (9)$$

T^2 = Treynor-squared measure

r_{p^*} = return on the adjusted portfolio

r_m = return on the market portfolio

The adjusted portfolio is the managed portfolio adjusted such that it has the same degree of systematic or market risk as the market portfolio. Since the market risk or beta of the market portfolio is equal to one, the adjusted portfolio is constructed as a combination of the managed portfolio and risk-free asset, such that the adjusted portfolio has a beta equal to one. The weights are specified as in Eqs. (10) and (11).

$$w_{rp} = \frac{\beta_m}{\beta_p} \quad (10)$$

$$w_{rf} = 1 - w_{rp} \quad (11)$$

w_{rp} represents the weight given to the managed portfolio, which is equal to the beta of the market portfolio (β_m) divided by the beta of the managed portfolio (β_p). w_{rf} is the weight on the risk-free asset and is equal to one minus the weight on the managed portfolio. The beta of the adjusted portfolio (β_{p^*}) is the weight on the managed portfolio times, the beta of the managed portfolio, and this will be equal to the risk of the market portfolio as shown in Eq. (12).

$$\beta_{p^*} = w_{rp} \times \beta_p = \frac{\beta_m}{\beta_p} \times \beta_p = \beta_m \quad (12)$$

The return of the adjusted portfolio (r_{p^*}) is computed as the weighted average of the returns of the managed portfolio and the risk-free rate (Eq. 13), where the weights are as determined above in Eqs. (10) and (11).

$$r_{p*} = w_{rf} \times r_f + w_{rp} \times r_p \quad (13)$$

The return on the adjusted portfolio can be readily compared with the return on the market portfolio since both have the same level of market risk. The differential return, T^2 , indicates the excess return of the managed portfolio in comparison to the benchmark portfolio after adjusting for differences in the market risk.

In the example, the beta of the managed portfolio is 1.5. Hence, the $w_{rp} = 1.0/1.5 = 0.67$, and $w_{rf} = 1 - 0.67 = 0.33$. The adjusted portfolio would be 67% invested in the managed portfolio and 33% invested in Treasury bills. The beta of the adjusted portfolio, $\sigma_{p*} = 0.67 \times 1.5 = 1.00$, which is equal to the beta of the market portfolio. The return on the adjusted portfolio would be $r_{p*} = 0.33 \times 4\% + 0.67 \times 20\% = 14.72\%$. The $T^2 = 14.72\% - 13.00\% = 1.72\%$. Thus, after adjusting for market risk, the managed portfolio has performed better than the benchmark by 1.72%. T^2 is a better measure for relative performance when the market risk of a managed portfolio is the relevant risk metric.

References

- Chen, Son-Nan, and C.F. Lee. 1981. The sampling relationship between Sharpe's performance measure and its risk proxy: Sample size, investment horizon and market conditions. *Management Science* 27: 607–618.
- . 1986. The effects of the sample size, the investment horizon and market conditions on the validity of composite performance measures: A generalization. *Management Science* 32: 1410–1421.
- Christopherson, Jon A. 1995. Equity style classifications. *Journal of Portfolio Management* 21: 32–43.
- Cumby, R.E., and J.D. Glen. 1990. Evaluating the performance of international mutual funds. *Journal of Finance* 45: 497–521.
- Grinblatt, M., and S. Titman. 1994. A study of monthly mutual fund returns and performance evaluation techniques. *Journal of Financial and Quantitative Analysis* 29: 419–444.
- Jensen, M.C. 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23: 389–416.
- Kallaberg, J.G., C.L. Lin, and C. Trzcinka. 2000. The value added from investment managers: An examination of funds of REITs. *Journal of Financial and Quantitative Analysis* 35: 387–408.
- Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–47.
- Modigliani, F., and L. Modigliani. 1997. Risk-adjusted performance: How to measure it and why. *Journal of Portfolio Management* 23: 45–54.
- Mossin, J. 1966. Equilibrium in the capital market. *Econometrica* 34: 768–783.
- Reilly, F.K., and E.A. Norton. 2003. *Investments*. 6th ed. Mason: Thompson-Southwestern.
- Sharpe, W.F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- . 1966. Mutual fund performance. *Journal of Business* 39 (1): 119–138.
- . 1992. Asset allocation: Management style and performance measurement. *Journal of Portfolio Management* 18: 7–19.
- . 1998. Morningstar's risk-adjusted ratings. *Financial Analysts Journal* 54: 21–33.
- Treynor, J.L. 1965. How to rate management of investment funds. *Harvard Business Review* 43: 63–75.

Robert A. Schwartz and Reto Francioni

Abstract

A call auction is an order-driven facility which, in contrast with continuous trading, batches multiple orders together for simultaneous execution in a multilateral trade, at a single price, at a predetermined point in time, by a predetermined matching algorithm. The chapter describes how orders are handled and clearing prices set in call auction trading, contrasts call auctions with continuous trading, and identifies different types of call auctions (including price scan auctions, sealed bid auctions, and open limit order book auctions). Attention is given to the use of information technology in call market design, the integration of an auction in a market's micro-structure, and to the facility's ability to deal with market quality issues such as containing intra-day price volatility, sharpening price discovery, and catering to participant demands for immediacy. To produce robust results, a call auction must attract sufficient critical mass order flow; the paper concludes by noting that, because large traders in particular are reluctant to enter their orders early in the auction process, book building cannot be taken for granted.

Keywords

Book building · Continuous trading · Critical mass order flow · Hybrid markets · Information technology · Intra-day price volatility · Open limit order book auction · Order-driven facility · Price and time priority · Price discovery · Price improvement · Price scan auction · Sealed bid auction

R. A. Schwartz (✉)

Baruch College of the City University of New York, New York, NY, USA

Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

e-mail: Robert_Schwartz@baruch.cuny.edu

R. Francioni

Swiss Stock Exchange, Zurich, Switzerland

e-mail: reto.francioni@unibas.ch

See note 1.

A call auction is an order-driven facility that batches multiple orders together for simultaneous execution in a multilateral trade, at a single price, at a predetermined point in time. This contrasts with continuous trading where a trade can occur whenever a buy and a sell order cross in price. Our discussion of call auction trading is implicitly in the context of equity trading, but the concepts involved apply to a far greater array of financial instruments and nonfinancial resources.

The call auction form of trading died out in the precomputer age but has made its re-entrance today as an electronic marketplace. An electronic call auction has been incorporated in recent years in a number of market centers around the world, most notably Deutsche Börse, Euronext (the Paris, Amsterdam, Brussels, and Lisbon exchanges), the London Stock Exchange, and the NASDAQ Stock Market. These electronic calls are not being used as standalone systems, but have been combined with continuous trading to create hybrid markets. When it comes to trading, one size does not fit all. With a hybrid system, an investor can select among alternative trading venues depending on the size of the order, the liquidity of the stock being traded, and the investor's own motive for trading.

A pure "order-driven" market is a trading environment where all of the participants are investors seeking to buy or to sell shares for their own portfolio purposes. The environment is called "order driven" because limit orders that are placed by some participants set the prices at which others can trade by market order. Most order-driven markets are not "pure," but allow for market making. Even without imposing specific obligations or offering incentives, large market participants often choose to make markets as a specific business line. There is a need for market making services, and these services get paid for.

An order-driven market can be structured in two fundamentally different ways. With a "continuous" market, a trade can be made at any moment in continuous time that a buy order and a sell order meet in price. In the continuous market, trading is generally a sequence of bilateral matches. Alternatively, in a "call auction," orders are batched together for a simultaneous execution. At the time of a call, a market clearing price is determined, and buy orders at this price and higher execute, as do sell orders at this price and lower.

Call auctions and continuous trading both have their advantages and their shortcomings. In most exchanges, both methods are combined, as are order-driven and quote-driven facilities¹ so as to form an optimum structure for all kinds of users. In principle, an auction appears to be the ideal way of defining the equilibrium market

¹Adapted from Robert A. Schwartz and Reto Francioni (2004), *Equity Markets in Action: The Fundamentals of Market Structure and Trading*, John Wiley (Copyright © 2004 Robert A. Schwartz and Reto Francioni; This material is used by permission of John Wiley); and Robert A. Schwartz, (2003) "The Call Auction Alternative," In Robert A. Schwartz, John Aidan Byrne and Antoinette Colaninno (eds.) *Call Auction Trading: New Answers to Old Questions*, Kluwer Academic Publishers (Springer).

price at a specific point in time. Continuous trading is more apt to resemble a never ending crawl around a dynamically evolving equilibrium price.

Many retail customers are accustomed to trading with immediacy. Nevertheless, if there were only retail orders, periodic calls would probably be the better way to provide fair and equitable treatment to every investor. However, markets must also cope with the problem of handling large block orders. A lot of interaction with the market is needed to trade large orders. That is where some see the advantage of continuous trading. It offers a special kind of interaction between the market participants, opportunities to test the market, and to get information from the market. For big orders, periodic calls may not provide the kind of flexibility that some participants want.

On both sides of the Atlantic, this has led to combinations of the call and continuous systems. Call auctions are typically used at the beginning of each trading session to open their continuous order-driven markets. The opening price has special importance because orders that have come in during the overnight trading halt are normally considered to have an equal right to get filled, at least partly, at the opening price. Setting the opening price should, therefore, be done carefully – be it by a well-structured auction or through a less formalized process. Calls are also used to close the market. The major European equity markets and the New York Stock Exchange and NASDAQ in the United States do this to sharpen the accuracy of price determination at this critical time of the trading day, and in recognition of the multiplicity of uses to which the closing prices are put (at index rebalancings and derivative expirations, as well as for marking-to-market in derivative markets, share valuations for various other legal purposes, etc.). Some exchanges also run periodic calls during a trading session (Deutsche Börse's market model includes one intra-day call). The intra-day calls are particularly important for securities with low trading volume.

36.1 Order Handling

Orders are handled differently in call auctions than in continuous trading, and the time clock is used differently. With a call auction, trades are made at specific points in time rather than whenever, in continuous time, a buy and a sell order cross. To accomplish this, orders submitted to a call auction that could otherwise have been matched and executed are batched together for a multilateral clearing. The clearings are generally held at predetermined points in time (at the open, at the close, and/or at set times during the trading day).

As noted, at the time of a call, the batched orders are matched, and a single clearing price is established. The single clearing price reflects the full set of orders submitted to the call. Buy orders at this value and higher execute, as do sell orders at this value and lower. Because all executed orders clear at the same price, there is no bid–ask spread in call auction trading. Further, with single price clearing, buy orders priced above the single clearing value and sell orders priced below it receive price improvement.

36.2 Alternative Call Auction Designs

Many variations in auction design exist. Calls can be held “on request” instead of at predetermined regular intervals. The amount of precall pricing information to reveal is a decision variable. Traders may be free to change their orders/quotes until the last moment, or there may be restrictions of various kinds. And so forth.

Taking an aerial view, we identify four basic types of call auctions (with several variations in between).

36.2.1 Price Scan Auctions

In a price scan auction, a sequence of prices is “called out” until a value is found that best balances the buy and sell orders.² The Paris Stock Exchange’s market, before the Bourse introduced electronic trading in 1986, was a classic price scan call auction. When the market for a stock was called, an auctioneer would cry out one price after another, scanning the range of possibilities, until an acceptable balance was found between the buy and sell orders.

36.2.2 Sealed Bid Auctions

In a sealed bid auction, participants submit their orders in sealed envelopes that are not opened until the time of the auction. These auctions are totally “closed book” (nontransparent) during the preopen phase, and consequently no participant knows what orders the others are submitting. The term may also be applied more broadly when orders are submitted electronically or by other means if pre-trade orders and indicated clearing prices are not revealed to participants. The US Treasury’s new issues market is a good example of the sealed bid auction.

In an electronic trading environment, the auction can be set up with various degrees of preauction transparency that allows traders to react to an indicated clearing price that is continuously displayed as the market forms. This functionality characterizes the next category of call auctions, open limit order book calls.

36.2.3 Open Limit Order Book

With an open limit order book, posted orders are displayed to the public in the precall order entry period. As the time of the call approaches, the procedure also identifies and updates an indicated clearing price which, at each instant, is the value that would be set in the call if the call were to be held at that instant. At the time of the call, the

²In a quote-driven market, the quotes of a dealer or market maker establish the prices at which others can trade by market order.

book is frozen and the indicated clearing price becomes the actual clearing price. Variants of the open limit order book call is used in most electronic order-driven trading platforms around the world.

The next category is not, strictly speaking, a call because it does not undertake price discovery. However, because it is based on the principle of order batching, we include it here: crossing networks,

36.2.4 Crossing Networks

A crossing network does not discover price. Rather, buy and sell orders are matched in a multilateral trade at a price that is set elsewhere. Generally, the price used at a cross is either the last transaction price or the midpoint of the bid–ask spread set in a major market center. In the United States, ITG’s cross is a good example of this facility.

36.3 Order Batching and Price Determination

Figures 1, 2, 3, and 4 describe order batching and price determination in a call. In each of these figures, share price is shown on the vertical axis, and the number of orders is shown on the horizontal axis. The number of shares sought for purchase or offered for sale is conventionally displayed on the horizontal axis, but the exposition is simplified by assuming that all orders are for the same number of shares (e.g., one round lot). The following legend is used in the diagrams:

- Individual buy order
- Cumulative buy orders at the price or better
- Individual sell order

○ Cumulative sell order at the price or better

Figure 1 displays the individual buy and sell orders. The horizontal axis gives the total number of orders (buys plus sells) that have been placed at each price. At each price, the orders are arrayed according to the sequence in which they have arrived.

Fig. 1 Batching of customer orders

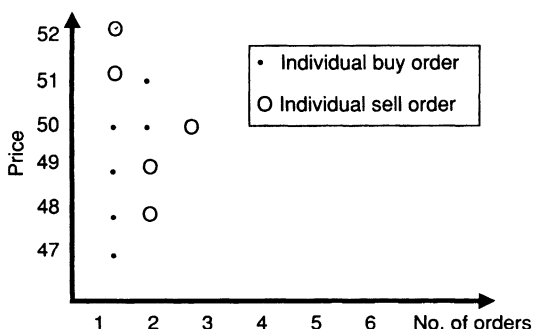
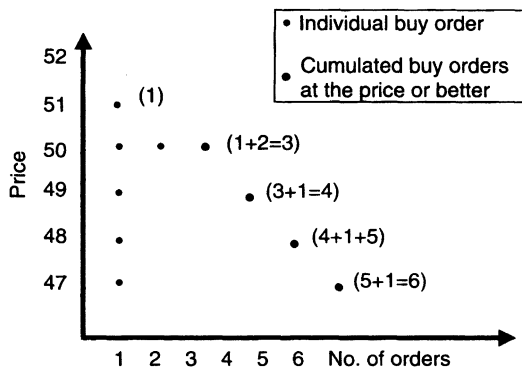
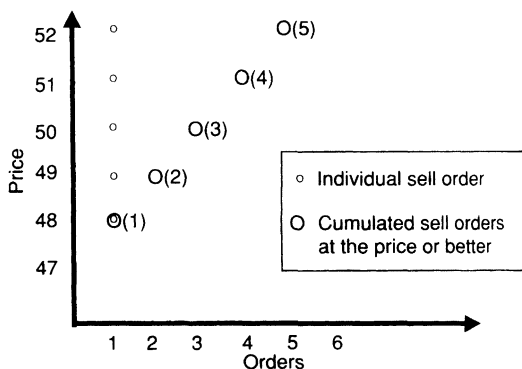


Fig. 2 Cumulation of the buy orders**Fig. 3** Cumulation of the sell orders

At the price of 52, just one sell order has been placed. At 51, a sell order arrived first, and then a buy order. At 50, two buy orders arrived followed by one sell order. And so on.

Figures 2 and 3 show how the individual buy and sell orders are aggregated. The buy orders only (both individual and aggregated) are shown in Fig. 2. Because the price limit on a buy order is the highest price at which the order is to be executed, the buy orders are cumulated from the highest price (in this case 51) down to the lowest (47). At 51, there is just one order to buy. Two additional buy orders have been entered at 50, and thus at 50, there is a total of three buy orders. At yet lower prices, one order has been placed at each of the prices, 49, 48, and 47. Thus, the cumulative number of orders at these prices is four, five, and six, respectively.

The sell orders only (both individual and aggregated) are shown in Fig. 3, and they are also cumulated. Because the price limit on a sell order is the lowest price at which the order is to be executed, the sell orders are cumulated from the lowest price (48) up to the highest price (52). There is only one sell order at each of the prices, and the cumulative number of sell orders increases by one order as we move from the single order at 48 to the five orders at 52.

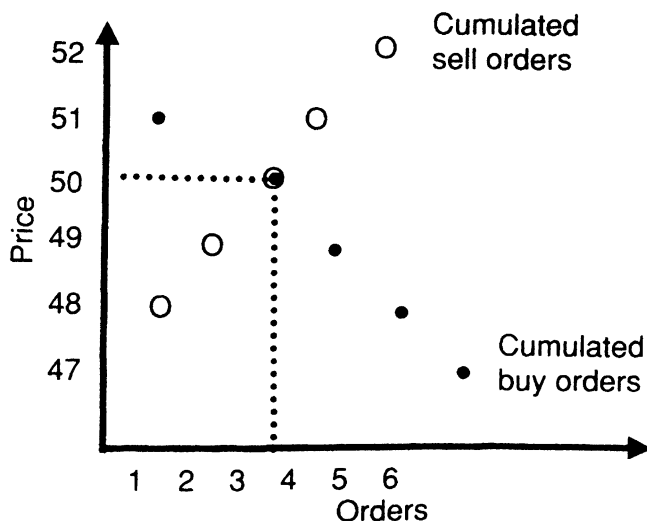


Fig. 4 Matching of the cumulated buy and sell orders

The cumulative buy and sell orders are matched together in Fig. 4 to determine the clearing price at which they execute and the specific orders that execute. At the intersection of the two curves, price is 50 and the number of orders is three. Thus, three buy orders execute (the one placed at 51 and the two at 50) and three sell orders execute (the one placed at 48, the one at 49, and the one at 50). Note that three is the maximum number of orders that can execute: at the higher price of 51 there is only one buy order, and at the lower price of 49 there are only two sell orders. For this reason, the clearing price in a call auction is typically identified as the value that maximizes the number of shares that execute (and, in the special case presented here, the number of orders that execute).

Note that the most aggressive buy orders are matched with the most aggressive sell orders. This is because orders receive price priority. Namely, the most aggressive orders (on either side) are executed first. As we discuss below, if several orders have the same price limits, the order that was input first gets executed first (time priority). In the example depicted in Fig. 4, three of the executed orders receive price improvement (the buy at 51, the sell at 49, and the sell at 48). The less aggressive orders (the buys at 49, 48, and 47, and the sells at 51 and 52) remain unexecuted. These orders may be rolled into the continuous market, held for the next call, or cancelled, depending on the wishes of the investor.

In Fig. 4, at the market-clearing price of 50, the cumulated sell orders match the cumulated buy orders exactly. What if no price exists that gives an exact match? For instance, what would happen if, everything else constant, three buy orders rather than two were entered at 50? The decision rule would still pick 50 to be the price (this value would still maximize the number of orders that execute), but with a cumulative of only three sell orders at 50, only three of the four buy orders can be executed.

A further decision rule is needed to specify which three of the four orders to pick. The rule commonly used is the “time priority rule:” orders execute according to the sequence in which they were placed, with the first to arrive being the first to execute. Time priority is valuable in call auction trading as it gives participants an incentive to place their orders earlier in the precall, order entry period.³

36.4 Relationship Between Limit and Market Orders

Limit orders and market orders are very different order types in continuous trading, but are virtually the same in call auction trading. For continuous markets, limit orders set the prices at which market orders execute, and limit orders sitting on the book provide immediacy to the market orders (i.e., the market orders execute upon arrival). Limit order traders are willing to wait patiently for an execution and they are the liquidity providers. In a continuous market, market order traders demand immediate liquidity.

In contrast, market orders in the call environment are nothing more than extremely aggressively priced limit orders. Specifically, a market order to buy has an effective price limit of infinity and a market order to sell has an effective price limit of zero. Participants in a call auction all wait until the next call for their orders to execute, and thus market orders in a call auction do not receive immediacy as they do in continuous trading. The distinction in continuous trading that limit order placers supply liquidity while market order placers demand liquidity, does not apply to call auction trading. In a call auction, all participants supply liquidity to each other. However, with an open book call, those participants who place their orders early in the precall order entry period are key to the book building process. As we discuss further below, early order placers are the catalysts for liquidity supply.

36.5 The Electronic Call Auction

One and a half centuries ago, the New York Stock Exchange was a call market (nonelectronic, of course). In some respects, the nonelectronic call was a fine system for participants on the exchange floor but it had serious deficiencies for anybody away from the floor. Investors not physically present had little knowledge of what was happening (the calls offered no transparency), and access to trading was limited because shares of a stock could be exchanged only periodically (when the market for the stock was called). On May 8, 1869, the call procedure was abandoned when the NYSE merged with a competing exchange, the Open Board of Brokers, and became a continuous trading environment.

³Further situations can be described that require more complex rules of order execution. As is typically the case, the set of decision rules required for an actual operating system is far more complicated than those we need consider to achieve a basic understanding of a system.

The Tel-Aviv Stock Exchange through the 1970s and the Paris Bourse before the 1986 introduction of its electronic market, CAC (the acronym stands for “*Cotation Assistée en Continu*”) were also nonelectronic call auctions that did not survive.

Call auction trading had been very popular with continental European exchanges in the earlier days when they still had floor trading. But with growing competition among exchanges, continuous trading became increasingly popular. The widespread trend to fully automated trading of most European exchanges, however, has allowed for new solutions and combinations.

Tremendous advances in information technology and a slew of other developments in the industry have paved the way for the call’s reentry. With an electronic open limit order book, participants anywhere around the globe are able to see the auction as it forms, and can enter their own orders with electronic speed. Compared to traditional floor trading, electronic trading offers new flexibilities for fine-tuning market architecture. Automated order book trading usually starts with an opening call and uses a call to resume trading after any halt. As noted, the major European exchanges, the New York Stock Exchange, and NASDAQ have also introduced closing electronic calls, particularly to provide “better” closing prices. In 2019, average daily trading volume in the NYSE’s closing call auction amounted to 9% of daily trading volume in the continuous market (see Ozenbas and Schwartz (2021)).

While information technology (IT) can be used advantageously in continuous trading, it is essential for efficient call auction trading. Moreover, the call auction is an extremely good environment for the application of IT. In a continuous market, IT speeds up the rate at which orders can be submitted, displayed, and turned into trades and, in so doing, it accentuates the importance of nanoseconds. In an electronic call auction environment, on the other hand, IT is used to sort and cumulate orders, and to find the clearing prices. In a call auction, the computer is used to do one thing in particular that it was created to do, namely, to compute.

The electronic call auction is appealing for small and mid-cap stocks because order batching augments the efficiency of liquidity provision by focusing liquidity at specific points in time. The procedure also has particular appeal for the large cap stocks, because it caters to the needs of institutional participants whose portfolios are mostly comprised of these issues. Market impact is reduced for the institutional investor because the call is a point in time meeting place and, as noted, batching orders in a multilateral trade focuses liquidity. For all stocks, commissions may be lower due to the greater ease of handling orders and clearing trades in the call auction environment.

For the broad market, electronic call auctions can reduce short-period (e.g., intra-day) price volatility, unreliable pricing, unequal access to the market, and various forms of manipulation and abuse.⁴ Further, the electronic call auction is an explicit

⁴For further discussion of the properties of call auction trading, see Cohen and Schwartz (1989), Economides and Schwartz (1995), Pagano and Schwartz (2003), Schwartz et al. (2006), Chap. 4, and Pagano et al. (2013).

price discovery facility. That is, batching many orders together for simultaneous execution at a single price produces a consensus value that better reflects the broad market's desire to hold shares. Consequently, the electronic call auction is a good opening facility for the continuous order-driven market. Moreover, because it is an explicit price discovery facility, call auction trading can be used to dampen short-period (e.g., intra-day) price volatility.

One feature of call auction trading that has been thought by some to be a drawback is that it does not provide transactional immediacy (participants have to wait for a call). With call and continuous trading combined in a hybrid market structure, this limitation ceases to be a deficiency. And, in any event, immediacy involves a cost (bid–ask spreads and market impact costs) that not all investors wish to pay. Retail and institutional customers who place limit orders are not looking for immediate executions and many institutional customers are more concerned with anonymity and keeping trading costs low than with obtaining immediate executions.

To deliver its promise of being a highly efficient trading environment, a call auction must attract sufficient volume. To accomplish this, some order placers must be incented to enter their orders early in the precall order entry period. The early stages of book building cannot be taken for granted, however, especially for an auction that opens the market at the start of a trading day. Some participants, particularly big institutional customers, are reluctant to post orders, an act that may reveal their trading intentions when the book is thin. Nevertheless, early order placers, the catalysts for liquidity supply, are needed. Two incentives for early order placement are (1) the use of time priorities and (2) reduced commission rates for early order entry. The inclusion of retail customers who are less concerned that their small orders will have any meaningful impact on the clearing price also helps. Lastly, a market maker could play an important role in animating book building during the precall order entry period.

References

- Cohen, K.J., and R.A. Schwartz. 1989. An electronic call market: Its design and desirability. In *The challenge of information technology for the securities markets: Liquidity, volatility, and global trading*, ed. H. Lucas and R.A. Schwartz, 15–58. Homewood: Dow Jones-Irwin.
- Economides, N., and R.A. Schwartz. 1995. Electronic call market trading. *Journal of Portfolio Management* 22: 10–18.
- Ozenbas, Deniz, and Robert A. Schwartz. 2021. *The Return of the Call Auction*. Working paper.
- Pagano, Michael, and Robert A. Schwartz. 2003. A closing call's impact on market quality at Euronext Paris. *Journal of Financial Economics* 68: 439–484.
- Pagano, Michael S., Lin Peng, and Robert A. Schwartz. 2013. A call Auction's impact on price formation and order routing: Evidence from the Nasdaq stock market. *Journal of Financial Markets*, May 2013.
- Schwartz, R.A., R. Francioni, and B.W. Weber. 2006. Decision making in equity trading using simulation to get a grip. *Journal of Trading* 1: 59–74.



Robert A. Schwartz and Lin Peng

Abstract

Liquidity, which is integrally related to trading costs, refers to the ability of individuals to trade at reasonable prices with reasonable speed. As such, liquidity is a major determinant, along with risk and return, of a company's share value. Unfortunately, an operational, generally accepted measure of liquidity does not exist. This chapter considers the following proxy measures: the bid-ask spread, the liquidity ratio (which relates the number or value of shares traded during a brief interval to the absolute value of the percentage price change over the interval), the variance ratio (which relates the volatility of short-term price movements to longer-term price movements), and resiliency. The determinants of liquidity considered are the size of the market for a stock and market structure. The entry concludes by stressing that illiquidity increases the cost of equity capital for firms, but that trading costs can be reduced and liquidity enhanced by the institution of a superior trading system.

Keywords

Bid-ask spread · Cost of equity capital · Liquidity · Liquidity ratio · Market structure · Risk and return · Share value · Trading costs · Trading system · Variance ratio · Resiliency

R. A. Schwartz (✉)

Baruch College of the City University of New York, New York, NY, USA

Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

e-mail: robert.schwartz@baruch.cuny.edu

L. Peng

Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

e-mail: lin.peng@baruch.cuny.edu

Liquidity refers broadly to the ability of individuals to trade quickly at prices that are reasonable in light of underlying demand/supply conditions. Liquidity, risk, and return are the major determinants of a company's share value. Risk constant, expected return must be higher and a company's cost of capital greater, if the market for its shares is less liquid. A number of authors have studied the cross-sectional relationship between liquidity and asset prices (see, e.g., Amihud and Mendelson 1986; Brennan and Subrahmanyam 1996; Easley et al. 2002; and Pastor and Stambaugh 2003), as well as the time series relationship (Jones 2002). However, a comprehensive understanding of the impact and determinants of liquidity is still lacking. The problem is that an operational, generally accepted measure of liquidity does not exist.

Liquidity is often described by the depth, breadth, and resiliency of the market for an asset. A market has depth and breadth if orders exist at an array of prices in the close neighborhood above and below the values at which shares are currently trading, and if the buy and sell orders exist in substantial volume. A market is resilient if temporary price changes due to order imbalances quickly attract new orders that restore reasonable share values.

Liquidity (and its converse, illiquidity) can also be defined in terms of the transaction costs incurred to obtain a fast execution. Transaction costs include an explicit component such as commissions and an implicit component such as a bid-ask spread and market impact. The ask quotation is the price at which shares can be purchased with immediacy, and the bid is the price at which shares can be sold with immediacy. The difference, known as the bid-ask spread, is the cost of a round-trip, and half of the spread is typically viewed as the cost of buying or selling shares immediately.

Market impact exists when a buy order drives the ask up, or a sell order drives the bid down. This occurs because the volume of shares at the quotes may be small relative to the size of the order, and/or because of the dissemination of the information that a large trader has arrived in the market. The spread and market impact are large if a market lacks depth and breadth.

Bid-ask spreads are directly quantifiable, but market impact is very difficult. The problem is two-fold. First, because of information leaks and front-running, an order can impact prices before it reaches the market. Second, prices are constantly changing due to news and liquidity trading, and thus a reasonable benchmark against which to assess the implicit cost components of a transaction price is not readily available.

Prices are also distorted due to the difficulty of finding equilibrium values in the marketplace. Errors in price discovery occur because prices depend on the order flow while simultaneously orders are priced with imperfect information about the underlying consensus values. Analogous to the market impact effect, transaction prices can be pushed up if impatient buyers outnumber impatient sellers, or can be pushed down if impatient sellers outnumber impatient buyers (Ho et al. 1985). In a resilient market, errors in price discovery are quickly corrected.

None of the attributes of liquidity thus far discussed provide an unambiguous measure of the concept. One commonly used measure is the bid-ask spread (Amihud and Mendelson 1986). Another is the liquidity ratio, which relates the number or value of shares traded during a brief time interval to the absolute value of the percentage price change over the interval. The larger the ratio of shares traded to

the percentage price change, the more liquid the market is presumed to be. This view underlies measures of price impact that have been used by various stock exchanges, and characterizes the approach taken by some researchers to measure and to contrast the liquidity of different market centers (Cooper et al. 1985; Hui and Heubel 1984).

The liquidity ratio, however, can be misleading. If news causes prices to change, a large liquidity ratio that is attributed to heavy trading volume would suggest that prices have adjusted too slowly in response to the informational change. This is because a bid that is too high attracts market orders to sell, and an ask that is too low attracts market orders to buy. Consequently, to the extent that trading is triggered by informational change (rather than by idiosyncratic investor needs), trading volume is less, and the liquidity ratio is smaller (not larger), in a more efficient market.

Another measure of liquidity is the variance ratio, which relates the volatility of short-term price movements to the volatility of longer-term price movements. Transaction prices jump up and down as executions bounce between the bid and the ask, as large orders impact prices, and as transaction prices fluctuate around equilibrium values due to price discovery errors. Thus, implicit execution costs increase the volatility of short-term price movements. Because the effect attenuates as the interval over which price changes are measured is lengthened, it is possible to proxy liquidity by the variance ratio. Hasbrouck and Schwartz (1988), for example, find that an appropriately adjusted ratio of two-day to half-hour returns variance is predominantly less than unity (the value expected for a perfectly liquid market) for a large sample of NYSE, Amex, and OTC stocks. Ozenbas, Schwartz, and Wood (2002) report an accentuation of intra-day volatility that is most pronounced in the first half-hour of a trading day in five markets – the New York Stock Exchange and Nasdaq in the USA, and the London Stock Exchange, Euronext Paris, and Deutsche Börse in Europe.

A primary determinant of liquidity is the size of the market for a stock (or, inversely, thinness). Size can be measured as the number or value of shares outstanding, the number or value of shares traded, and/or the number of shareholders. Empirical studies have shown that spreads are wider, market impact greater, and price discovery less accurate for thinner issues (Cohen et al. 1986; Schwartz and Francioni 2004). But even for larger issues, markets can be thin, particularly for big, institutional investors. This is because, during any trading session, only a relatively small number of individuals actually seek to trade. For small-cap and mid-cap stocks, the problem may be particularly striking within a trading day: at any given moment, only a handful of individuals (if any) may be actively looking to buy or to sell shares.

Hua, Peng, Schwartz, and Alan (2020) present an expanded definition of liquidity, resiliency, that encompasses both the price impact of a liquidity shock and its persistence. They further show that stock resiliency is priced and that higher resiliency is robustly associated with lower expected returns. Crucially, resiliency matters for returns even for the large cap stocks and those in the S&P 500 index, suggesting that liquidity is a key factor in asset pricing across a broad spectrum of stocks. Market structure also affects the liquidity of individual issues, and the US Securities and Exchange Commission has required that execution venues report their execution quality on multiple dimensions (see SEC 2000). The primary market makers in certain market centers are dealers and exchange market makers (formally

known as “specialist”), whose role is to supply immediacy to public traders. In this context, the provision of immediacy is essentially synonymous with the provision of liquidity, the ability to transact quickly at reasonable prices. Liquidity may also be enhanced by other market-structure mechanisms. One important approach would be to increase the depth and breadth of a market by encouraging public traders to place more limit orders. The imposition of rules to prevent destabilizing trades (i.e., tick-test rules) and the time batching of orders in call auction trading are two other ways to increase liquidity. In 2001, the NYSE and Nasdaq completed a conversion from fractional to decimal prices under pressure from the SEC. The switch has resulted in sharply reduced quoted spreads. However, there is evidence that the inside market depth has been reduced for the large traders (Sofianos 2001).

Public orders generally execute at inferior prices in illiquid markets. As a consequence, expected returns on securities traded in less liquid markets must be higher, and the cost of capital for the listed companies is greater. The important insight is that the costs of trading can be decreased by the institution of a superior trading system. In the limit, as a market becomes frictionless, the issues traded in it become perfectly liquid.

References

- Amihud, Yakov, and Haim Mendelson. 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17 (2): 223–249.
- Beebower, Gilbert, V. Kamath, and R. Surz. 1985. Commission and transaction costs of stock market trading. Working Paper, SEI Corporation, July.
- Bernstein, Peter. 1987. Liquidity, stock markets, and market makers. *Financial Management* 16: 54–62.
- Brennan, Michael J., and Avanidhar Subrahmanyam. 1996. Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics* 41 (3): 441–464.
- Cohen, Kalman, Steven Maier, Robert A. Schwartz, and David Whitcomb. 1986. *The microstructure of securities markets*. Englewood Cliffs: Prentice-Hall.
- Cooper, Kerry, John Groth, and William Avera. 1985. Liquidity, exchange listing, and common stock performance. *Journal of Economics and Business* 37 (1): 19–33.
- Easley, David, Soeren Hvidkjaer, and Maureen O’Hara. 2002. Is information risk a determinant of asset returns? *Journal of Finance* 57: 2185–2221.
- Economides, Nicholas, and Aloysius Siow. 1988. The division of markets is limited by the extent of liquidity. *American Economic Review* 78: 108–121.
- Grossman, Sanford J., and Merton Miller. 1988. Liquidity and market structure. *Journal of Finance* 43: 617–633.
- Hasbrouck, Joel, and Robert A. Schwartz. 1988. Liquidity and execution costs in equity markets. *Journal of Portfolio Management* 14 (3): 10–17.
- Ho, Thomas, Robert A. Schwartz, and David Whitcomb. 1985. The trading decision and market clearing under transaction price uncertainty. *Journal of Finance* 40 (1): 21–42.
- Hua, Jian, Lin Peng, Robert A. Schwartz, and Nazli Sila Alan. 2020. Resiliency and stock returns. *The Review of Financial Studies* 33 (2): 747–782.
- Hui, Baldwin and Heubel, Barbara. 1984. Comparative liquidity advantages among major U.S. stock markets. DRI Financial Information Group Study Series No.84081, 1984.

- Jones, Charles M. 2002. A century of stock market liquidity and trading costs, working paper, Columbia University.
- Lippman, Steven A., and John J. McCall. 1986. An operational measure of liquidity. *American Economic Review* 76 (1): 43–55.
- Ozenbas, Deniz, Robert A. Schwartz., and Robert. Wood. 2002. Volatility in U.S. and European equity markets: An assessment of market quality. *International Finance*, 437–461.
- Pastor, Lubos, and Robert F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111 (3): 642–685.
- Schwartz, Robert A., and Reto Francioni. 2004. *Equity markets in action: The fundamentals of market structure and trading*. Hoboken, New Jersey.
- Sofianos, George. 2001. *Trading and market structure research*. Hoboken, New Jersey: Goldman Sachs.
- U.S. Securities and Exchange Commission (SEC). 2000. Disclosure of order routing and execution practices, Release No. 34–43084, July 28.

Robert A. Schwartz and Lin Peng

Abstract

The primary focus of this chapter is on market maker services, revenues, and costs. A market maker's basic function is to service the public's demand to trade with immediacy by continuously standing ready to buy shares from customers who wish to sell, and to sell shares to customers who wish to buy. Additionally, the market maker helps to stabilize prices and to facilitate reasonably accurate price discovery. Further, a special type of market maker, a stock exchange market maker, fulfills the role of an auctioneer. The bid-ask spread is the classic source of market maker profits, while the costs of market maker operations include: order processing, risk bearing (the cost of carrying an unbalanced portfolio), and adverse selection (the cost of trading with a better informed participant). The entry further considers the competitive environment that market makers operate within, and concludes with the thought that institutionalization, the advent of electronic trading, and globalization of the equity markets have led to major changes in market maker operations in the recent past, and will continue to do so in the coming years.

This material is modified from an equivalent entry from: *The New Palgrave Dictionary of Money and Finance*, by: Newman, Peter. Reprinted with permission of Palgrave Macmillan. Copyright © Newman, Peter.

R. A. Schwartz (✉)

Baruch College of the City University of New York, New York, NY, USA

Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

e-mail: robert.schwartz@baruch.cuny.edu

L. Peng

Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

e-mail: lin.peng@baruch.cuny.edu

Keywords

Adverse selection · Bid-ask spread · Competitive environment · Electronic trading · Immediacy · Market maker · Order processing · Price stabilization · Price discovery · Risk bearing

Market makers play a central role in many equity markets by buying and selling shares to service the public's demand to trade immediately (the classic service provided by a dealer). Stock exchange market makers are also responsible for stabilizing prices (making a market "fair and orderly") and facilitating price determination. New York Stock Exchange market makers have also performed the role of auctioneer.

Demsetz (1968) was one of the first to analyze the supply of immediacy. Buyers and sellers arrive sporadically at the market, and it is not a simple matter for them to find each other in time. The market maker provides a solution by continuously standing ready to trade from his or her own inventory of shares. The service is not free, however. The dealer sells to buyers at higher ask prices, and buys from sellers at lower bid prices. The bid-ask spread is the market maker's compensation (sometimes referred to as "the dealer's turn").

Market makers are not necessary for immediacy to be provided to a market. Public traders can post limit orders with commission brokers acting as middlemen. However, immediacy is not the only marketability service provided by a market maker.

The liquidity provided by market makers also helps to stabilize prices. Most participants in the securities markets prefer prices that, all else equal, are less volatile. They care about this as investors because they are generally assumed to be risk averse. They care about this as traders because they are averse to transaction price uncertainty. Market maker intervention helps to stabilize price fluctuations in the short run. The US exchange market makers in particular have also had an "affirmative obligation" to make a fair and orderly market.

Market makers also facilitate the determination of accurate prices. First, their own quotes directly set market prices. Second, their quotes are signals that public traders react to in writing their orders; therefore, market makers indirectly affect market prices by influencing the public order flow. Third, exchange market makers establish market clearing prices at the opening of the trading day and at the resumption of trading after halts caused by the advent of news.

Price stabilization and price discovery are both consistent with the provision of immediacy. This is because "immediacy" means not only the ability to trade promptly, but also the ability to trade in reasonable amounts at prices that properly reflect current market conditions. (Smidt [1971] emphasizes the supply of liquidity in depth, namely the ability of investors to trade quickly and in size, at the market maker's quotes.) Consequently, transactional immediacy, price stability, and accurate price discovery are all attributes of markets that are "fair and orderly."

As auctioneers in an agency market, market makers also organize and oversee trading. Stock exchange market makers do so by maintaining the limit order books and by assuring that trading rules are not violated. On some exchanges (such as the Tokyo Stock Exchange), market makers act only in the clerical bookkeeping and

regulatory oversight capacities, and are not allowed to trade the stocks assigned to them.

The bid-ask spread set by the market maker reflects the following components: order-processing costs, risk premium or inventory costs, adverse selection costs, and profit (Stoll 1989). The order-processing costs compensate market makers for their time and effort, cost of paperwork, etc. Risk bearing is central to the dealership function (Amihud and Mendelson 1980; Ho and Stoll 1981). The market maker trades to make a market rather than for his or her own investment motives. If buyers appear, a market maker must be willing to assume a short position; if sellers arrive, he or she must be willing to assume a long position. As a result, the market maker generally acquires an unbalanced portfolio. He or she is then subject to uncertainty concerning the future price and the future transactions volume in the asset. Not knowing when transactions will be made, the market maker does not know for how long an unbalanced inventory position will have to be maintained. An unbalanced inventory position implies the existence of diversifiable risk. Thus the market maker requires a risk premium on the inventory risk which other investors can eliminate by proper portfolio diversification (the expected return on a stock compensates all investors and market makers for accepting nondiversifiable risk).

Market makers also protect themselves against adverse selection. Public orders to purchase or to sell securities are motivated by either idiosyncratic liquidity reasons or informational change. The market maker typically does not know whether an order has originated from an informed trader or from a liquidity trader. If a public trader receives news and transmits the order before the market maker has learned of the informational change, the public trader profits at the market maker's expense (Bagehot 1971; Copeland and Galai 1983; and Glosten and Milgrom 1985). The market maker responds to the cost of ignorance by increasing the ask quote and by lowering the bid so that the expected loss to the informed traders is compensated by the expected gain from the liquidity traders. He or she cannot achieve total protection, however, by sufficiently widening the spread. Regardless of how much the offer is raised and/or the bid is lowered, any informationally motivated trade would be at the market maker's expense.

And the defensive maneuver is not costless. The market maker profits from liquidity trades and, in the process of widening the spread to guard against informed traders, increases the cost of transacting and so loses an increasing number of liquidity traders. Yet, there must be investors who trade for noninformational reasons. Without the liquidity traders, the dealer market would collapse (Grossman and Stiglitz 1980).

The competitive environment of a market maker firm differs depending on whether it operates in an agency/auction environment or in a dealer market. In an agency/auction market, limit-order traders and floor traders provide competition for the exchange market makers (formally known as "specialist"). In contrast, a dealer market is competitive only if the order flow for a security is directed to more than one dealer firm. This competition for marketability services fragments the informational content of the order flow, however. That is, each dealer firm knows what buy and sell orders it receives, but does not observe the flow of orders to competing dealer firms. However, in a screen-based system each dealer firm does see the quotes posted by others. In addition, information is transferred by transaction price reporting and via inter-dealer trading.

The online reporting of large transactions, however, can signal information about a dealer's inventory position to its competitors. And, when the order flow is dominated by institutional investors, other issues can arise (see Neuberger and Schwartz 1990). These include fair-weather market making (taking the privileges but failing to meet the obligations of market making), preferencing (the diversion of order flow to a market maker firm that is not necessarily posting the best quotes, but that has guaranteed best price execution nonetheless), handling a lumpy order flow (few trades but of large size), and coping with one-way markets (buyers only or sellers only). All told, market making is a complex, multifaceted operation.

Institutionalization, the advent of electronic trading, and globalization of the equity markets are having a profound impact on securities trading and price determination. These forces have led to major changes in market maker operations in the recent past, and will continue to do so in the coming years.

References

- Amihud, Yakov, and Haim Mendelson. 1980. Dealership market: Market making with inventory. *Journal of Financial Economics* 8: 31–53.
- Amihud, Yakov, Thomas Ho, and Schwartz, Robert A. (eds). 1985. *Market making and the changing structure of the securities industry*. Lexington: Lexington Books.
- Bagehot, Walter. 1971. The only game in town. *Financial Analysts Journal* 27: 12–14, 22.
- Cohen, Kalman, Steven Maier, Robert A. Schwartz, and David Whitcomb. 1979. Market makers and the market spread: A review of recent literature. *Journal of Financial and Quantitative Analysis* 14: 813–835.
- Copeland, Thomas, and Daniel Galai. 1983. Information effects on the bid-ask spread. *Journal of Finance* 38 (December): 1457–1469.
- Demsetz, Harold. 1968. The cost of transacting. *Quarterly Journal of Economics* 82: 33–53.
- Garman, Mark. 1976. Market microstructure. *Journal of Financial Economics* 3: 257–275.
- Glosten, Lawrence, and Paul Milgrom. 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14: 71–100.
- Grossman, Sanford J., and Joseph Stiglitz. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70: 393–408.
- Ho, Thomas, and Hans Stoll. 1981. Optimal dealer pricing under transaction and return uncertainty. *Journal of Financial Economics* 9: 47–73.
- . 1983. The dynamics of dealer markets under competition. *Journal of Finance* 38: 1053–1074.
- Neuberger, Anthony, and Robert A. Schwartz. 1990. Current developments in the London equity market. *Finanzmarkt und Portfolio Management* 3: 281–301.
- O'Hara, Maureen, and George S. Oldfield. 1986. The microeconomics of market making. *Journal of Financial and Quantitative Analysis* 21: 361–376.
- Schwartz, Robert A., and Reto Francioni. 2004. *Equity markets in action: The fundamentals of market structure and trading*. Hoboken, New Jersey: Wiley.
- Smidt, Seymour. 1971. Which road to an efficient stock market: Free competition or regulated monopoly? *Financial Analysts Journal* 27 (18–20): 64–69.
- Stoll, Hans. 1985. Alternate views of market making. In Amihud, Ho and Schwartz (eds) *Market Making and the Changing Structure of the Securities Industry*. Lexington Books.
- . 1989. Inferring the components of the bid-ask spread: Theory and empirical tests. *Journal of Finance* 44: 115–134.
- Tinic, Seha M., and Richard West. 1972. Competition and the pricing of dealer services in the over-the-counter stock market. *Journal of Financial and Quantitative Analysis* 7: 1707–1727.



Robert A. Schwartz and Lin Peng

Abstract

The chapter reviews essential elements of market structure – the systems, procedures, and protocols that determine how orders are handled, translated into trades, and transaction prices determined. Various alternatives are contrasted such as: order-driven and quote-driven markets; consolidated vs. fragmented markets; human intervention vs. electronic trading; and continuous markets vs. periodic call auctions. A major objective of market design that is noted in the discussion is to enhance the accuracy with which prices are discovered in a dynamic, uncertain environment. Lastly, the chapter points out that market structures are rapidly changing, and that much remains to be learned about how best to structure a technologically sophisticated, hybrid market that efficiently services the varied needs of diverse participants.

Keywords

Call auctions · Consolidated markets · Continuous markets · Electronic trading · Fragmented markets · Hybrid market · Market structure · Order-driven markets · Price discovery · Quote-driven markets

This material is modified from an equivalent entry in Peter Newman, Murray Milgate and John Eatwell Editors, *The New Palgrave Dictionary of Money and Finance*, 1992, Published in the United Kingdom by The Macmillan Press Limited, London and Basingstoke, Published in the United States of America and Canada by The Stockton Press, New York. Reproduced with permission of Palgrave Macmillan.

R. A. Schwartz (✉)

Baruch College of the City University of New York, New York, NY, USA

Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

e-mail: robert.schwartz@baruch.cuny.edu

L. Peng

Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

e-mail: lin.peng@baruch.cuny.edu

The structure of a securities market refers to the systems, procedures, and protocols that determine how orders are handled, translated into trades, and transaction prices determined. To date, theoretical security valuation models have generally not considered the effect of a market's structure on asset prices. Formulations such as the capital asset pricing model and the arbitrage pricing theory, for example, address the risk and return dimensions of a security, but ignore considerations such as liquidity, trading costs, information costs, and transaction uncertainty. When these realities are taken into account, it is apparent that market structure matters, and that it does affect the price and size of trades.

Market structures differ significantly among major international equity market centers (see Schwartz and Francioni 2004). The New York Stock Exchange (NYSE) and other US stock exchanges are agency/auction markets where the market maker acts as both dealer and broker's broker. Examples of a dealer market include the Nasdaq market in the USA and the London Stock Exchange (LSE) before they introduced their electronic order-driven trading systems. The Tokyo Stock Exchange (TSE) is an agency/auction market where the market maker (*saitori*) handles the orders but does not take a dealership position. Markets also differ in the way in which orders are consolidated or fragmented, and in the way in which information is disseminated and displayed.

Whether investors trade through an intermediary (as in a dealer market) or directly with each other (as in an agency/auction market) is one of the most important distinctions in market structure. In a dealer market, the market maker initiates trades by posting bid and ask quotations that are publicly disseminated. The bid is the price at which public traders can sell to a dealer, and the ask is the price at which they can buy from a dealer. The bid-ask spread is the dealer's compensation for providing marketability services. To achieve a trade in a dealer market, a customer (usually via a broker) contacts a dealer by telephone or electronically and accepts his or her quotation.

In an agency/auction market, public participants trade with each other, and floor professionals in an agency market such as the NYSE act in a brokerage (agency) capacity. When trading is active in a stock, floor traders gather in a "crowd," and trading truly takes place in an auction environment. In the US exchanges, orders are consolidated at the posts of market makers (formally called "specialists"), who are professionals who function as both principals and as agents. Specialists have an affirmative obligation to buy and to sell shares so as to make "a fair and orderly market" when counterpart orders do not provide sufficient liquidity. They also have a negative obligation: when a public order and a market maker's quote are at the same price, the specialist must step aside and let the public order execute first.

Two types of orders are commonly used in an agency/auction market: limit orders and market orders. A limit order states the maximum price at which a public investor is willing to *buy*, or the minimum price at which the public investor is willing to *sell*, a specified number of shares. A market order is unpriced; it states the number of shares the investor wishes to trade "at market," namely the price prevailing when the order is received by the market center. For a market order to be executed, limit orders must exist; for limit orders to exist, there must be a facility for maintaining the public orders in a file (known as a limit order book). This file characterizes agency/auction

exchanges. Handa and Schwartz (1996) have examined the costs and returns to placing limit orders.

Trades may also be negotiated if they are difficult to handle because of their size. In an agency/auction environment such as the NYSE, a buyer or seller may give a not held (NH) order to a floor trader who uses his or her discretion to negotiate with other floor traders or to expose the order to the limit order book. The floor trader is “not held to the price” if the order executes at a price inferior to that which existed at the time of its arrival. Large orders are also negotiated in the “upstairs market,” a network of trading desks of securities dealers and institutional investors that bring buyers and sellers together at mutually acceptable prices. Trades may also be negotiated with a dealer and/or electronically through a facility such as Liquidnet or Coda Block. Institutional investors commonly negotiate with the market makers to obtain larger sizes than the market makers are quoting and/or prices that are within the bid-ask spread. Large orders are also commonly broken up (sliced and diced) and brought to the market in smaller tranches for execution over an extended period of time.

A major function of a market center is to find the prices at which shares are traded. This process is known as “price discovery.” The accuracy of price discovery depends on the systems used for handling orders, disseminating information, and making trades. If an issue is traded in more than one market center, intermarket linkages including information systems and arbitrage operations must be implemented to ensure both adequate price protection for investors and price consistency across markets. Intermarket linkages also connect equity markets and derivative product markets (for example, the futures and options markets for stock indices in Chicago, and the cash market for shares in New York).

Another feature of market structure is the means by which information concerning current market conditions (floor information) is transferred among participants. The informational signal transmitted by a quote differs significantly from that transmitted by a transaction price. A quote reflects one individual’s willingness to trade, is firm only up to its stated size, and may be improved on in terms of price and/or quantity. Quotes may also reflect trading strategy and gaming by market participants. A transaction price has actually been accepted by both counter parties to a trade, but relates to the past and does not necessarily represent the price at which one can trade in the present. Nonetheless, last transaction prices do reflect current market conditions when transactions occur frequently. For this reason, transaction price reporting has been introduced in both the USA and London dealer markets (see Seguin 1991).

The extent to which orders are fragmented or consolidated in trading also defines a market’s structure (see Cohen et al. 1986). A competitive dealer market is naturally fragmented in that orders are routed to one of several dealer firms. This may be desirable because of the competition for marketability services that fragmentation implies. Most apparent is that bid-ask spreads are tightened in a competitive dealer environment compared to a monopoly dealer environment (see Ho and Stoll 1983). However, this environment can also create incentives for dealers to collude (see Christie and Schultz 1994; Christie et al. 1994). In 1996, the justice department settled with the Nasdaq dealers on accusations of spread collusion.

Another problem of the dealer market is that fragmenting the order flow across different dealer firms can obscure information and impair the accuracy of price determination (see Neuberger and Schwartz 1990). However, in a screen-based system such as the US NASDAQ market, each dealer firm does see the quotes posted by the others. A dealer market with fragmented orders may also reduce the opportunity for the interaction of all buying and selling interest in that security and thus reduce price competition. In 1997, the US Securities and Exchange Commission enacted the order-handling rules (OHRs), which required that public limit orders be exposed in the national best bid and offer (NBBO). The rules set in motion the transition of the Nasdaq market from a predominantly quote-driven, dealer market toward an order-driven, agency market.

Order flow in an agency/auction environment is by its nature more consolidated than in a competitive dealer market. Consolidation is desirable because it allows orders to be matched against each other with a minimum of broker-dealer intervention. Furthermore, the consolidation of orders facilitates the enforcement of order exposure and trading priority rules. The primary priority rule is price; highest-priced bids and lowest-priced asks have precedence. A secondary priority rule specifies the sequence in which orders at the same price execute; usually, the first order entered at the price is the first to execute (time priority). However, too much consolidation may lead to monopoly power for a single market center, which may lead it to lose its incentive to reduce transaction costs and to innovate.

An agency/auction market is fragmented when shares are listed on more than one exchange, traded in house by a brokerage firm, on an ATS (alternative trading system) and/or on an ECN (electronic communications network). This fragmentation may be desirable if it truly represents competition between market centers. It is not desirable if one market center free rides on the prices discovered by another market center. For example, a satellite market may guarantee trades at the best price quoted in a major market center and charge lower commissions for the service.

Order consolidation facilitates the consolidation and transference of floor information. For example, NYSE market makers have been in a unique position from which to observe the order flow and to set prices that are reasonable given the current demand for shares. But, like the *saitori* in Tokyo, NYSE market makers are not permitted to receive orders directly from customers, which restricts their access to information. In contrast, dealers can receive orders directly from customers, including institutional traders. This contact enables them to obtain further information about market conditions.

In addition to being spatially (geographically) consolidated, orders can be consolidated temporally (over time). Orders are temporally consolidated when they are bunched together in call auction trading. In continuous trading, orders are executed whenever they cross during trading hours and, in a continuous market, trades are generally bilateral. In contrast, in call auction trading, orders are stored for simultaneous execution in multilateral trades at predetermined times when the market is "called."

Call market trading has certain advantages (see Schwartz and Francioni 2004). In particular, dependence on the intermediation of dealers and brokers is lessened

and trading costs are reduced. Because everyone trades at the same price, at the same time, and under the same conditions, call market trading is fairer, and the procedure can produce prices that are more accurate and less volatile. But, traditional call market trading has had limitations. Accessibility to the market was restricted and the dissemination of floor information poor in the old call markets of yesteryear. These limitations can be overcome with the use of computer technology. Pagano and Schwartz (2003) have found that the introduction of electronic call auctions at market closings on the Paris Bourse (now Euronext Paris) reduced transaction costs and improved price discovery. Pagano et al. (2013) find that the NASDAQ's introduction of electronic call auctions have had a positive spillover effect on price formations in the continuous market: the closing call significantly reduced both spreads and volatility for all market capitalization groups, and the opening call had a similar effect on spreads and a somewhat weaker effect for volatility.

One of the more striking changes in market structure that occurred as the twentieth century drew to a close was the advent of electronic trading. At its inception, electronic systems tended to mimic existing systems; now they are more commonly developing their own distinctive functionality. The first electronic exchange, the computer-assisted trading system (CATS), was introduced by the Toronto Stock Exchange in 1977. CATS is based on the principle of continuous trading in an agency/auction environment. The success of CATS has led to the implementation of similar systems in Tokyo (1982), Paris (1986), and elsewhere. Small-order execution systems were also introduced in the USA and London dealer markets in the 1980s. Now national equity markets around the globe provide floorless, electronic trading platforms.

Electronic technology has strong advantages: it gives participants direct access to markets and control over their orders regardless of geographic location; it provides direct access to information concerning current market conditions; it provides anonymity; it makes it possible for investors to trade without a broker and thus reduce transaction costs; and, as systems become increasingly sophisticated, the computational power of the computer has facilitated the handling of institutional-sized orders and the negotiation of trades. Investors in the 1990s have witnessed a proliferation of fourth market organizations. Electronic facilities such as Instinet and Archipelago allow members to post orders and to match the orders of other traders in the system. Crossing systems such as Posit and Instinet's crossing network allow investors to trade portfolios directly without a bid-ask spread. Liquidnet and Pipeline allow participants to find each other on their screen and to negotiate their trades electronically.

Electronic technology solves the major problems associated with nonelectronic call market trading: restricted accessibility to a market and inadequate dissemination of floor information (see Pagano and Schwartz 2003). Reciprocally, a call market environment may be more suitable than the continuous market for the use of electronic technology. In particular, the submission and handling of institutional-sized orders can be accommodated in an electronic call (see Schwartz and Francioni 2004).

Because of strong vested interests, technological inertia, and the ability of an established market center to retain order flow, the superiority of a new system may not ensure its acceptance. Market structure has evolved slowly in the USA since trading moved from coffee houses and curbs into exchanges. The pace of change accelerated in the mid-1970s with the passage of the US Securities Acts Amendments of 1975, which precluded fixed commissions and mandated the development of a national market system. London's Big Bang in 1986 also precluded fixed commissions, broadened competition between dealers and brokers, and further spurred the globalization of trading. More recently, the NYSE and Nasdaq have completed a conversion from fractional to decimal prices under the pressure of the SEC. Technological developments, inter-market competition, and regulation will no doubt continue to reshape securities markets around the world. However, achieving meaningful change in market structure is not an easy task, and much remains to be learned about how best to structure a technologically sophisticated, hybrid market that efficiently services the varied needs of diverse participants.

References

- Christie, W.G., and P.H. Schultz. 1994. Why do NASDAQ market makers avoid odd-eighth quotes? *Journal of Finance* 49: 1813–1840.
- Christie, W.G., G.H. Harris, and P.H. Schultz. 1994. Why did NASDAQ market makers stop avoiding odd-eighth quotes? *Journal of Finance* 49: 1841–1860.
- . 1986. *The microstructure of securities markets*. Englewood Cliffs: Prentice-Hall.
- Handa, P., and R. Schwartz. 1996. Limit order trading. *Journal of Finance* 51: 1835–1861.
- Ho, T., and H. Stoll. 1983. The dynamics of dealer markets under competition. *Journal of Finance* 38: 1053–1074.
- Neuberger, A., and R. Schwartz. 1990. Current developments in the London equity market. *Finanzmarkt und Portfolio Management* 3: 281–301.
- Pagano, M., and R. Schwartz. 2003. A closing call's impact on market quality at Euronext Paris. *Journal of Financial Economics* 68: 439–484.
- Pagano, M., L. Peng, and R. Schwartz. 2013. A call auction's impact on price formation and order routing: Evidence from the NASDAQ stock market. *Journal of Financial Markets* 16 (2): 331–361.
- Schwartz, R., and R. Francioni. 2004. *Equity markets in action: The fundamentals of market structure and trading*. Hoboken: Wiley.
- Seguin, P. 1991. Transactions reporting, liquidity and volatility: An empirical investigation of national market system listing. Paper given at Western finance association meetings, Jackson Hole, Wyoming, June.



Accounting Scandals and Implications for Directors: Lessons from Enron

40

Pearl Tan and Gillian Yeo

Abstract

We analyze the Enron case to identify the risk factors that potentially led to its collapse and specific issues relating to its aggressive accounting and high-light the lessons for independent directors. In Enron, the interactions between external stimuli, strategies, corporate culture, and risk exposures possibly created an explosive situation that eventually led to its demise. Much of the post-Enron reforms have been directed towards regulating the roles and responsibilities of executive directors and auditors. However, the role of independent directors has received relatively lesser attention. Independent directors should analyze the risks of their companies and understand the pressures that arise from market conditions and firm-specific policies and incentive structures. They also need to close the information gap between executive directors and themselves. A post-Enron era also requires independent directors to change their focus. Traditionally, independent directors have to strike a difficult balance between maximizing returns and minimizing risks. Independent directors may now have to focus on the management of risks, the design and functioning of an effective corporate governance infrastructure, and the moderation of the power bases of dominant executives. Practically, they may also have to reduce the number of independent director appointments to enable them to focus more effectively on a fewer companies.

P. Tan (✉)

Singapore Management University, Singapore, Singapore
e-mail: pearltan@smu.edu.sg

G. Yeo

Nanyang Technological University, Singapore, Singapore
e-mail: ahhyeo@ntu.edu.sg

Keywords

Accounting scandals · Audit committee · Corporate governance · Hedging · Independent directors · Incentives · Risks · Sarbanes–Oxley Act · Special purpose entity · Volatility

40.1 Introduction

The recent spate of accounting scandals raises serious concerns about the opportunistic use of accounting procedures and policies to camouflage fundamental problems in companies. The series of corporate collapses also highlight the failure of corporate governance mechanisms to prevent and detect accounting irregularities. The convergence of several factors, including competitive pressures, conflicts of interest, lack of market discipline, and inherent limitations of accounting standards resulted in an explosive situation whereby managers use aggressive accounting practices to present financial statements that do not reflect economic reality. In this chapter, we analyze the Enron case with the objective of determining the risk factors that potentially led to its collapse and specific issues relating to its aggressive accounting and highlight the lessons to be learnt for corporate governance from the perspective of an independent director.

40.2 The Competitive Environment and Incentives for Aggressive Accounting

Enron was formed as a result of merger of two companies in 1985. The merger was funded by debt and pressure had existed from the start for the new company to reduce its debt burden. At about the same time, deregulation of the natural energy industry exposed Enron to substantial operating and price risks arising from the increase in gas supply and volatility in spot prices. However, deregulation also increased opportunities for more flexible and innovative contracts to be drawn up between the producer and buyers. To survive, Enron had to capitalize on these opportunities and became a primary market player through its development of the idea of a Gas Bank. Under this scheme, Enron facilitated the market for energy contracts by buying gas from suppliers and selling to buyers. In acting as an intermediary, Enron guaranteed both the supply and the price and assumed the related risks in return for transaction fees. Innovations were subsequently extended to markets for basic metals, pulp and paper, and broadband products. Its diversification strategy also included investments in other countries in South America, Europe, and Asia. The business and geographical diversification created new risks for Enron. Its heavy investment in projects such as broadband network assets would pay off only in the long term. However, an immediate debt burden from these

acquisitions placed pressure on Enron's balance sheet that was already weighed down by existing debt (Powers et al. 2002).¹

Although Enron began as an operator of energy-related assets, by the end of the 1990s, the firm had divested a significant portion of its physical assets in what is known as an "asset light strategy" (Permanent Subcommittee on Investigations of the Committee of Governmental Affairs 2002)² and was primarily focused on its trading and financial activities relating to physical energy commodities. Effectively, the company was transformed from a natural gas supplier into an energy trader and intermediary. It offered specialist services in price risk management strategies and market-making activities. Its dominance in the market for energy contracts gave Enron a first-mover advantage in exploiting information economies of scale. However, the lucrative profits it enjoyed attracted other entrants to the industry and Enron's profit margins began to erode by the end of 2000. Further, as a trader, Enron was compelled to maintain an investment grade rating in order to lower its counter-party risk.

Against this backdrop of competitive pressures, Enron's senior management developed incentive schemes that turned the firm environment into a highly competitive internal market place. An internal ranking system administered by the company's Performance Review Committee became a means of allocating bonus points and determining dismissals. The entire process was described as a "blood sport" (Chaffin and Fidler 2002) and former employees believed that the basis for reward was largely determined by whether a deal could be reported as revenue or earnings rather than commitment to the company's core values of Respect, Integrity, Communication, and Excellence. Enron's annual incentive awards and the long-term incentive grants are closely tied to company performance measures and stock prices. The annual incentive bonus was pegged to a percentage of recurring after-tax profit, while its long-term incentive grants provided for accelerated vesting provided Enron achieved performance targets linked to compounded growth in earnings-per-share and cumulative shareholder returns.³ A Senate report on the Enron collapse concluded that Enron's Board of Directors approved lavish and excessive executive compensation and failed to stem the "cumulative cash drain" arising from its incentive schemes.⁴

Hence, Enron appeared to react to risk by creating an environment that generated new risk exposures through its business strategies and reward system that focused on short-term results. Figure 1 summarizes the competitive pressures at Enron.

¹Hereinafter referred to as the "Powers Report."

²Hereinafter referred to as the "Senate Report," p. 7.

³Proxy Statement Pursuant to Section 14(a) of the Securities Exchange Act of 1934, 2 March 2001, EDGARPlus(R).

⁴The Senate Report, p. 3.

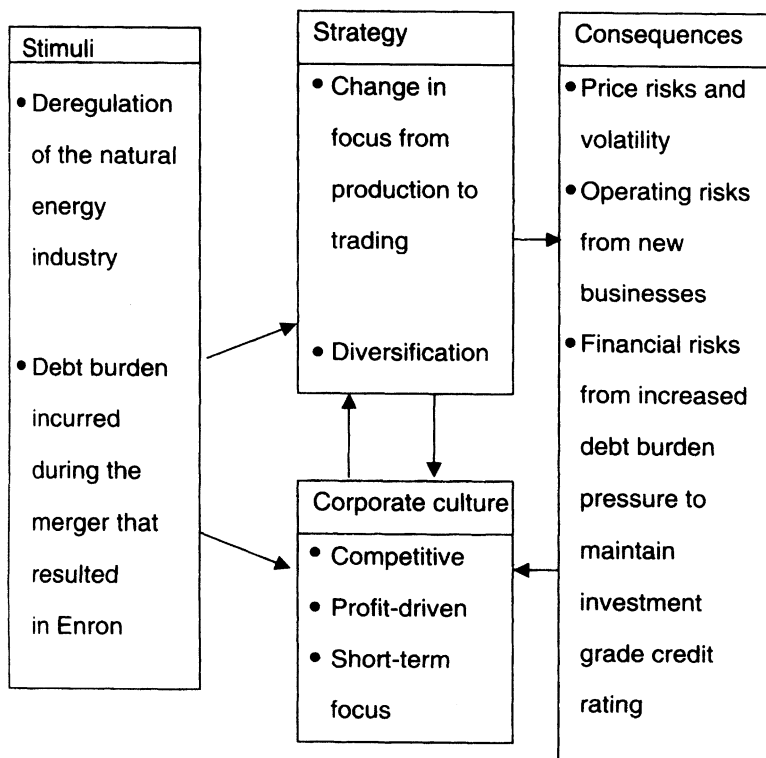


Fig. 1 Competitive pressures at Enron

40.3 Aggressive Accounting Practices

Enron's accounting practices resulted in removing the liabilities of its balance sheet, improving profitability, and reducing profit volatility. These desired accounting effects were achieved through structuring numerous complex and "innovative" transactions. Many of these transactions involved dealing with special purpose entities that Enron set up in partnership with related parties. The investigating Senate Committee described these practices as "high-risk accounting." The manner in which certain transactions were reported was deemed to be at variance with their true economic substance. The main question that underlies these practices relates to the issue of whether Enron had retained the risks that were purportedly transferred to the special purpose entities.

40.3.1 Effectiveness of "Hedging" Transactions

An example included the entering into transactions that were purported to hedge the volatility of its "marked to market" investments. The hedging transactions were

entered into between Enron and a special purpose entity (SPE).⁵ A hedge is effective only if a loss suffered by a hedged party is transferred out to an outside party. In its first hedging transaction, Enron transferred its own stock to the SPE in exchange for a note. The intention of the hedge was to transfer losses to the SPE, through the exercise of an option, should the stock price of a profitable “merchant” investment decline. The SPE purported to take on the risk of price volatility of the investment and to compensate Enron for the loss on its investments.

However, cash was available to the SPE only if the latter sold the Enron stock. Since the SPE was financed by Enron’s stock, the transaction was effectively a self-hedging arrangement as the creditworthiness of the SPE was tied to Enron’s fortunes. When Enron’s stock fell in value in late 2000 and early 2001, the SPE faced a liquidity crisis and could not honor its obligations under the option. Hence, the “hedge” was ineffective because the counter-party’s risk was inextricably intertwined with Enron’s risk and the hedge did not constitute a true economic hedge.

40.3.2 Control and Risks Relating to Unconsolidated Entities

There are two broad approaches in accounting for an SPE. If an SPE is controlled by an investing company, the assets and liabilities of the SPE are consolidated entirely on to the investing company’s balance sheet. Alternatively, if it is not under the investing company’s control, it is treated as an investment in a separate entity, with off-balance sheet treatment of the SPE’s assets and liabilities. Under applicable accounting rules in the United States at that time, an SPE could receive off-balance-sheet treatment only if independent third-party investors contributed at least 3 % of the SPE’s capital. Some of Enron’s dealings raised serious questions about whether this rule was effectively met.

For example, from 1997 to 2001, Enron did not consolidate an SPE called Chewco. In 1997, Enron and the California Public Employees’ Retirement System (CalPERS) were joint venture partners in an off-balance sheet investment vehicle called Joint Energy Development Limited Partnership (JEDI). To enable CalPERS to cash out its investment in JEDI in order to invest in a larger Enron venture, Andrew Fastow, the then Chief Financial Officer at Enron, and others at Enron formed an SPE called Chewco to buy CalPERS’ interest in JEDI. Thus, Enron was able to continue accounting for JEDI as an off-balance-sheet entity on the basis that the holdings by Enron staff members and related parties constitute outside capital at risk. According to SEC investigations,⁶ Fastow secretly controlled Chewco. Hence, a serious question arose as to whether Enron, through a related party, had effective control over major operating and financial policies of Chewco. Further, Enron and its related SPEs provided guarantees and cash collateral on bank funding to Chewco, indicating that equity at risk was effectively borne by Enron rather than independent

⁵Details of the hedging transactions are found in The Powers’ Report, pp. 13–15.

⁶Securities and Exchange Commission, Litigation Release 17762, 2 October 2002.

third parties. In November 2001, both Enron and its auditors, Arthur Andersen, concluded that Chewco was an SPE without sufficient outside equity and should have been consolidated. The retroactive consolidation of Chewco from 1997 to 2001 had an astounding effect on the financial statements. Profits decreased by a total of \$405 million over the period of restatement and additional debt of \$711 million was recognized on the balance sheet in 1997.⁷

40.4 The Role of Corporate Governance

Theoretically, Enron had in place an impressive array of corporate governance mechanisms. Outside directors were well-respected and highly qualified individuals in the fields of accounting, finance, and law. The Board of Directors had several committees to review various aspects of the company's policies and operations. There was separation of the offices of the Chairman and Chief Executive Officer. The external auditors were a then Big Five accounting firm. However, following the company's massive financial collapse, serious doubts arose as to the effectiveness of these institutional arrangements. The Senate Investigating Committee found that the Enron's Board failed to safeguard Enron shareholders and contributed to the collapse of the company by allowing Enron to engage in high-risk accounting, inappropriate conflict of interest transactions, extensive undisclosed off-balance-sheet activities, and excessive executive compensation.⁸ Further, the Board was also found to have failed to ensure the independence of the company's external auditor, Andersen who provided internal audit and consulting services as well.⁹

Many valuable lessons can be learnt from the Enron case to prevent the derailing of the effective functioning of governance mechanisms. We focus our discussion on the role of independent directors. Much of the post-Enron reforms have been directed towards regulating the roles and responsibilities of executive directors and auditors. However, the role of independent directors has received relatively less attention than that of other corporate governance agents. We discuss below some implications of the Enron collapse on the role of independent directors.

1. What is the primary role of independent directors? The multiple roles that independent directors have to undertake require them to strike a difficult balance between maximizing returns and minimizing risks. Their purview is wide, ranging from activities that have a "profit" focus to others that have a "defensive" focus. Independent directors potentially find themselves in an identity crisis. For example, if an independent director has to operate within an Enron-type environment, the director is confronted with an aggressive risk-taking internal

⁷The Powers' Report, p. 42.

⁸The Senate Report, p. 11.

⁹The Senate Report, p. 54.

environment. The question arises as to whether the independent director should act as a thorn in the managers' flesh or go with the flow of an aggressive managerial style for the sake of profit maximization?

The lesson from Enron is very clear that it does not pay to sacrifice the defensive role when risk factors are overwhelming and the long-run survival of the company is at stake. While post-Enron legislation such as the Sarbanes–Oxley Act of 2002 is primarily directed towards establishing mandates for insiders, audit committee board members, and external auditors, much less is said about the responsibilities of independent directors per se. However, the implicit responsibilities of independent directors are clearly reinforced by laws that impose fiduciary duties on directors to act in good faith, with reasonable care, and in the best interest of the corporation and its shareholders. The Conference Board also reiterates directors' role to monitor management and to ensure their ethical and legal compliance (The Conference Board 2003).¹⁰

Hence, independent directors owe a primary duty of care to outside investors. Their priority should be towards establishing and ensuring a corporate environment and infrastructure wherein managerial stewardship is executed without compromising the long-run interests of the firm and its stakeholders. They, more than anyone else, are best placed to limit the excesses of a dominant Chief Executive.

2. Independent directors have to bridge the information gap between executive directors and themselves. The Conference Board emphasizes that directors need to understand, among other things, the business strategies they approve, the risks and vulnerabilities arising from the strategies, growth opportunities, debt levels, and company's capital allocation of the companies under their purview.¹¹ Following the Enron experience, independent directors are well advised to understand the internal dynamics, managerial incentives, and power bases within the corporate environment and to adopt a healthy skepticism of strategies that potentially advance managerial interests over that of external investors. They should be keenly aware of the threats posed by dominant Chief Executive Officers and key personnel and the risks of opportunistic managerial behavior.
3. Greater commitment in terms of time and effort are expected of independent directors to meet the governance objective. Independent directors must take a proactive role in governance and not rely solely on external auditors, legal counsel, or key executives to provide them the necessary assurance. For example, when the Enron Board was asked why they moved so quickly in their approval of an unusual hedging transaction, the response was that the company had obtained a fairness opinion from an outside accounting firm.¹² On another proposal, the Board relied on the company's legal counsel to advise if anything was amiss on a

¹⁰ Hereinafter referred to as The Conference Board Report.

¹¹ The Conference Board Report, p. 9.

¹² The Senate Report, p. 27.

particular memorandum. Had the directors reviewed the memorandum for themselves, they would have noted that key company executives were involved in the arrangement that gave rise to conflicts of interest.¹³ Interviewed Board members told the investigating Senate Subcommittee members that they assumed that the then Chief Executive Officer had actively reviewed and approved the fairness of the unusual business proposals and the compensation controls.¹⁴ Enron's directors were also found to have knowingly allowed Enron's use of "high-risk" accounting without enforcing restraint.¹⁵ Hence, the Senate Report underscores the principle that evidence of a suspect transaction or activity that is known to a director must be questioned and examined diligently and thoroughly, regardless of the views of other experts.

The implications for independent directors are enormous. The days when an independent director held several of such appointments concurrently are likely to be over. Independent directors may have to be selective in choosing appointments so as not to spread themselves too thinly. They must also be prepared to commit resources and time and change the mindset that their appointment is a "part-time" one. They may also have to assess the risks of companies to determine if they are willing to undertake the fiduciary responsibility of monitoring such a company.

40.5 Conclusion

The Enron case has painful lessons for the business community. A seemingly successful company was apparently derailed through the use of highly risky transactions and aggressive accounting that temporarily boosted profits and reduced debt. The question arises as to why the corporate guardians of Enron did not prevent these transactions from occurring. Following Enron and other accounting scandals, a re-examination needs to be carried out of the role and responsibilities of independent directors. This chapter suggests that significantly greater challenges are posed to independent directors in a post-Enron world to understand more of the risks, accounting practices, and managerial opportunism existing in the companies under their purview and to take a more proactive role in governance, which inevitably requires a substantial commitment of their time and resources.

¹³The Senate Report, p. 28.

¹⁴The Senate Report, pp. 30–31.

¹⁵The Senate Report, pp. 14–24.

References

- Chaffin, J., and S. Fidler. 2002. The Enron collapse. *Financial Times*, London, 9: 30.
- Permanent Subcommittee on Investigations of the Committee of Governmental Affairs. 2002. United States Senate, The Role of the Board of Directors in Enron's Collapse Report, 107–170.
- Powers, Jr. W.C., R.S. Toubh, and H.R. Winokur, Jr. 2002. Report of investigation by the special investigative committee of the board of directors of Enron Corp.
- Securities and Exchange Commission. 2002. Litigation Release No. 17762.
- The Conference Board. 2003. Commission on public trust and private enterprise.



Nicholas S. P. Tay

Abstract

This chapter introduces the agent-based modeling methodology and points out the strengths of this method over traditional analytical methods of neoclassical economics. In addition, the various design issues that will be encountered in the design of an agent-based financial market are discussed.

Keywords

Agent-based models · Artificial intelligence · Bounded rationality · Classifiers · Co-evolution · Complex adaptive system · Computer simulation · Genetic Algorithms · Genetic programming · Heterogeneous agents · Learning · Neural networks

41.1 Introduction

The sort of phenomena that are interesting in finance and yet difficult to investigate analytically involve the complex interactions among many self-interested heterogeneous boundedly rational agents acting within the constraints imposed by either formal or informal institutions or authorities. To outrival their opponents, each and every agent must continually evolve to adapt to changes that may arise either from exogenous perturbations to the environment or endogenous transitions caused by agents changing their strategies or modifying their behaviors as they learn more about the behaviors of the other agents and the environment they reside in. A good example of such complex adaptive systems is the stock market.

A natural way to study a complex adaptive system like the stock market is to use an agent-based model which entails simulating the stock market on a computer from

N. S. P. Tay (✉)

University of San Francisco, San Francisco, CA, USA

e-mail: tay@usfca.edu

the bottom up with a large number of interacting heterogeneous boundedly rational artificial agents that are created to mimic the traders in the stock market. Once the environment of the stock market and the behaviors of the agents are specified and the initial state of the model is set, the dynamics of the model from the initial state forward will be driven entirely by agent-agent interactions, and not by some exogenously determined systems of equations. Hence, if any macroscopic regularity emerges from the model, it must be a product of the endogenous repeated local interactions of the autonomous agents and the overall institutional constraints. This is the spirit of the agent-based modeling approach.

What makes the agent-based modeling methodology particularly appealing? To begin with, analytical tractability is not an issue since this approach relies on computer simulations to understand the complex model. Quite the reverse, it is inconceivable how one could obtain closed form solutions of a model as complex as the stock market without first diluting drastically the authenticity of the model. Although analytically tractable heterogeneous agent rational expectations models have been around, the complexity and realism that are captured in agent-based models are beyond the reach of those analytical models.

For instance, consider the problem that a decision maker faces when the outcome is contingent on the decisions to be made by all the participating heterogeneous decision makers, each with their own unique preferences and quirks and private information that are not directly observable by the other decision makers. This decision problem is inherently ill defined and cannot be solved through mathematical deduction or analytical modeling. In real life, when confronted with such an ill-defined situation, decision makers often rely on the rules of thumb that they have distilled from years and years of experience to guide them in their decision-making. This decision making process is formally known as inductive reasoning and it can be captured naturally with the agent-based approach by running computer simulations of a large number of interacting artificial agents who make decisions using rules of thumb that they distill from their repeated interactions with each other.

The ability to build more realistic models with the agent-based method often allows agent-based models to reveal a much richer set of behaviors that are embedded in a system which may otherwise be overlooked by traditional equation-based models. For instance, Parunak et al. (1998) in comparing the differences between equation-based modeling and agent-based modeling of a supply network have found that equation-based model fails to produce many of the rich effects, such as memory effect of backlogged orders, transition effects, or the amplification of order variation, which are observed in an agent-based model of the same supply network. In addition, various agent-based models (Farmer and Joshi 2000; Johnson et al. 2001; LeBaron et al. 1999; Tay and Linn 2001) have been successful in accounting for real financial markets phenomena such as market crashes, mean reversion, relatively high level of trading, technical trading, excess volatility, and volatility clustering. These are phenomena that analytical representative agent models of financial markets have tolled to explain without much success.

Another serious shortcoming of analytical representative agent models of financial markets is that by design these models do not specify the dynamic process that

will need to happen in order to arrive at the equilibrium or equilibria that are characterized in these models. Consequently, for models that produce multiple equilibria, it is unclear which equilibrium among the multiple equilibria agents would converge on. In contrast, the events that unfold in a computer simulation of an agent-based model are completely transparent, and can be recorded hence providing the modeler a means to go back in the time line of evolution to understand how certain equilibrium or other global regularities came into existence.

The agent-based methodology therefore offers important advantages over the traditional analytical tools of neoclassical economics as it allows a researcher to obtain more germane results. Needless to say, the use of computer simulations as a tool for studying complex models has only become feasible in recent years because of the availability of fast and cheap computing power. Although the agent-based modeling methodology is still in its infancy, there is already a considerable literature on agent-based models. Leigh Tesfatsion at the Iowa State University maintains a website at <http://www.econ.iastate.edu/tesfatsi/ace.htm> to facilitate access to the extensive resources related to the agent-based modeling methodology, and to keep researchers in this field abreast of the latest developments.

In the introductory remarks on her website, Tesfatsion observes that agent-based research may generally be organized according to one of the following four research objectives: (1) empirical understanding, (2) normative understanding, (3) qualitative insight and theory generation, and (4) methodological advancement. The first objective focuses on seeking answers that are established on the repeated interactions of agents to explain the emergence of global regularities in agent-based models. Some examples of global regularities in financial markets are mean reversion and volatility clustering. Researchers in this group are interested in understanding if certain types of observed global regularities can be attributed to certain types of agent-based worlds. The second objective concerns using agent-based models as laboratories to aid in the discovery and design of good economic policies or good institutional structures. Researchers with this objective in mind are interested in using agent-based models to evaluate whether certain economic policies or institutional designs and processes will promote socially desirable outcomes over time among agents that are driven solely by their self interests. Tesfatsion phrased the third objective as “How can the full potentiality of economic systems be better understood through a better understanding of their complete phase portraits (equilibria plus basins of attraction)?” Unlike analytical models, the causal mechanisms in agent-based models are not direct and are very difficult to discern because of the complex nature of the interactions among the agents and between the agents and the environment. The goal here is to use the phase portraits as a means to enrich our understanding of the causal mechanism in these systems. The fourth objective addresses issues related to improving the methods and tools used by agent-based researchers.

For someone who is just starting out in this line of research, it is worthwhile to begin by reading “A Guide for Newcomers to Agent-based Modeling in the Social Sciences” by Axelrod and Tesfatsion which is available on the homepage of Tesfatsion’s website. In addition, it is beneficial to read the survey articles written by Hommes (2006), Duffy (2006), LeBaron et al. (1999), LeBaron (2000, 2006), and

Tesfatsion (2002) and a book by Batten (2000) that provides an overview of agent-based models and offers some historical perspectives of this methodology.

The next section discusses the design issues that will be encountered in the design of an agent-based model. This discussion benefited greatly from the insights that LeBaron has provided in his excellent overviews of the various design issues (LeBaron, 2000, 2001c, 2004a).

41.2 Design Considerations

A typical agent-based model is made up of a set of autonomous agents that encapsulate the behaviors of the various individuals in a system we are interested in studying and the investigation involves simulating on a computer the interactions of these agents over time. Accordingly, there are two important design considerations in the development of an agent-based model – the design of the agents and the design of the environment.

How naive or sophisticated the agents should be modeled really depends on the objective of the research. For instance, if the research objective is to understand how certain market structures affect the allocative efficiency of a market independent of the intelligence of the agents as in Gode and Sunder (1993), then one can simply model the agents as naive “zero intelligence” agents. Zero intelligence agents are agents that are not capable of formulating strategies or learning from their experience; hence their behaviors will be completely random. Gode and Sunder populated their double auction market with zero intelligence agents that are designed to submit their bids and asks at random over a predefined range and remarkably they discover that zero intelligence agents when subjected to a budget constraint are able to allocate the assets in the market at over 97% efficiency. The lesson to be learned here is that not all macroscopic regularities that emerge from agent-based models are necessarily consequences of the actions taken by the agents as they evolve and learn from their interactions. In this case, the high level of allocative efficiency that is attained in a double auction market is due to the unique structure of the market itself.

However, in many agent-based models, the objective is to investigate the outcome of the interactions among many heterogeneous agents that are designed to mimic their counterparts in the real world. In these models, the key design issues related to the design of the agents are the agents’ preferences and their decision-making behaviors. Agents could have either myopic or intertemporal preferences. The latter is more realistic but will make the model much more complex. As we have alluded to earlier, the decision problem that the agents face is usually ill defined, and thus cannot be solved by deductive reasoning. A reasonable solution is to assume that the agents rely on inductive reasoning to arrive at a decision (see Arthur 1994, 1999; Rescher 1980). Inductive reasoning or induction is a means for finding the best available answers to questions that transcend the information at hand. In real life, we often have to draw conclusions based upon incomplete information. In these instances, logical deduction fails because the information we have in hand leaves gaps in our reasoning. In order to complete our reasoning, we fill those gaps in the

least risky, minimally problematic way, as determined by plausible best-fit considerations. Consequently, the conclusions we draw using induction are suggested by the data at hand rather than logically deduced from them.

Inductive reasoning follows a two-step process: possibility elaboration and possibility reduction. The first step involves creating a spectrum of plausible alternatives based on our experience and the information available. In the second step, these alternatives are tested to see how well they answer the question at hand or how well they connect the existing incomplete premises to explain the data observed. The alternative offering the “best fit” is then accepted as a viable explanation. Subsequently, when new information becomes available or when the underlying premises change, the fit of the current alternative may degrade. When this happens a better alternative will take over.

How can inductive reasoning be implemented in an agent-based financial market model? Arthur (1994, 1999) envisions inductive reasoning in a financial market, taking place as follows. Initially, each agent in the market creates a multitude of decision-making rules (this corresponds to the possibility elaboration step discussed above). Next, the decision-making rules are simultaneously tested for their effectiveness based on some criteria. Finally, effective decision-making rules are retained and acted upon in buying and selling decisions. Conversely, unreliable rules are dropped (this corresponds to the possibility reduction step). The rules that are dropped are then replaced with new ones in the first step and the process is carried out repeatedly to model how individuals learn inductively in a constantly evolving financial market.

Some examples of criteria that have been used for appraising the effectiveness of the decision rules includes utility maximization, wealth maximization, and forecast errors minimization. Once a decision has been made on a criterion for evaluating the decision-making rules, the next task is to decide the length of historical data to be used in computing the criterion. Although many agent-based models tend to allow the agents to adopt identical history length, this is not necessary. It is in fact more realistic to permit agents in the same model to adopt different history length as in LeBaron (2001a, b).

To take the modeling to the next step, decision will have to be made concerning what the decision making rules look like and how they are to be generated in the models? One possibility is to model the decision-making rules after actual trading strategies used in real financial markets. The benefit of this approach is that the results are likely to be tractable and precise and it will also shed light on the interaction among these actual trading strategies. However, this approach does not allow the agents any flexibility in modifying the strategies or developing new strategies. This could impose ad hoc restrictions on the model’s dynamics. Some common tools that have been employed to allow the agents more degrees of freedom in structuring and manipulating the decision making rules as they learn are artificial neural networks (LeBaron2001a), genetic programming (Chen and Yeh 2001), and classifiers that are evolved with genetic algorithms (LeBaron et al. 1999). Even with these artificial intelligence tools, the modeler will need to predefine a set of information variables and functional forms to be used in the strategies or decision-making

rules. Although these tools can successfully mimic the inductive reasoning process described earlier, it is not known if any of these tools indeed faithfully represent the inductive reasoning used by actual human traders. It is also unclear at this juncture whether this issue matters. Another related decision that has to be made by the modeler concerns whether the agents should be allowed to learn only from their own experiences or from the collective experience of all the agents in the model. The latter is known as “social learning.”

We will turn our attention next to the design of the financial market environment. Most agent-based models simplify the environment to a market with one risky asset and one risk-free asset. Clearly, this is an oversimplification of actual financial markets, but there are good reasons for doing so. Given that the agent-based methodology is new and researchers barely comprehend the implications of this methodology, it is prudent for them to begin by exploring what the new method can reveal about the dynamics in a fairly simple market environment. Moreover, doing so also facilitates comparisons with results from well-known neoclassical models of a market with one risky asset and one risk-free asset.

Another key design issue concerns the design of the trading mechanism that has a direct influence on how prices are determined in the market and how the market is cleared. LeBaron (2004a) observes that most of the agent-based models employ one of the following four designs for trading mechanism. The simplest trading mechanism is one that allows mutually beneficial trades to be consummated between agents that meet at random. Though this trading mechanism is quite simple, it bears some resemblance to the trades conducted on the floor of the Chicago futures and options exchanges and over the telephone in the foreign exchange markets. But for markets where the market makers play an important role in filling the buy and sell orders, this mechanism would not be an adequate representation. A second trading mechanism, which is more sophisticated than the previous one is an analytical market-clearing device akin to one espoused in Grossman (1976). This device provides a closed form solution for the market-clearing price hence enabling the agent-based markets to be cleared analytically each period. A critical advantage of this design is that it avoids having to deal with the difficult issue of explicitly modeling the decision-making behaviors of the risk-adverse market maker. Unfortunately, this advantage is also a serious shortcoming in that this is not a realistic picture of what is happening in real markets that trade continuously and are rarely in equilibrium. The third trading mechanism attempts to address this issue. It assumes that agents submit trade orders to buy (D_t) and sell (S_t) at a price, p_t , which is announced beforehand by a market maker. The market maker then appraises the aggregate of the orders submitted by the agents, and adjusts next period price by a fixed fraction of the excess demand or supply according to $p_{t+1} = p_t + \alpha(D(p_t) - S(p_t))$. Granted that this adaptive price process may be a more reasonable model of how prices adjust in real markets, the problem with this mechanism is that it does not address how the market maker manages the imbalance between demand and supply in the market. Moreover, there is no guidance on how the parameter value for α should be determined and certain α values may in fact cause the market to deviate far from the market clearing price for a substantial period. The most sophisticated and

also the most realistic trading mechanism is one that either models the market maker explicitly or implements an order book system that can accept and cross out the buy and sell orders from agents according to some defined procedure (Audet et al. 2001). The only downside of this approach is that the design of the agent-based model is much more complicated as many details at the institutional as well as the agent level will need to be clearly specified. But, this is inevitable if the objective is to simulate realistic market microstructure behavior.

To sum up, there are many design questions that need to be addressed in the development of an agent-based model and there is yet no clear guidance on how best to address these questions. Inevitably, design decisions will have to be made however arbitrary these decisions may be but it is important to keep in mind that the choices made by the designer may ultimately have important consequences on the results.

References

- Arthur, W.B. 1994. Inductive reasoning and bounded rationality. *American Economic Review* 84 (2): 406.
- . 1999. Complexity and the economy. *Science* 284 (5411): 107.
- Audet, N., Gravelle, T., and Yang, J. 2001. Optimal market structure: Does one shoe fit all? Technical report, Ottawa: Bank of Canada.
- Batten, D.F. 2000. *Discovering artificial economics: How agents learn and economies evolve*. Boulder: Westview Press.
- Chen, S.-H., and C.-H. Yeh. 2001. Evolving traders and the business school with genetic programming: A new architecture of the agent-based artificial stock market. *Journal of Economic Dynamics and Control* 25: 363–394.
- Duffy, J. 2006. Agent-based models and human subject experiments. In *Handbook of computational economics volume 2: Agent-based computational economics*, Handbooks in Economics Series, ed. K.L. Judd and L. Tesfatsion. Amsterdam: Elsevier North-Holland.
- Farmer, J.D., and Joshi, S. 2000. The price dynamics of common trading strategies. Santa Fe Institute Working Paper 00–12–069, Santa Fe, New Mexico.
- Gode, D.K., and S. Sunder. 1993. Allocative efficiency of markets with zero intelligence traders. *Journal of Political Economy* 101: 119–137.
- Grossman, S.J. 1976. On the efficiency of competitive stock markets where traders have diverse information. *Journal of Finance* 31 (2): 573.
- Hommes, C.H. 2006. Heterogeneous agent models in economics and finance. In *Handbook of computational economics volume 2: Agent-based computational economics*, Handbooks in economics series, ed. K.L. Judd and L. Tesfatsion. Amsterdam: Elsevier North-Holland.
- Johnson, N.F., D. Lamper, P. Jeffries, M.L. Hart, and S. Howison. 2001. Application of multiagent games to the prediction of financial time-series. *Physica A* 299: 222–227.
- LeBaron, B., W.B. Arthur, and R. Palmer. 1999. Time series properties of an artificial stock market. *Journal of Economic Dynamics and Control* 23 (9, 10): 1487–1516.
- LeBaron, B. 2000. Agent-based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control* 24 (5–7): 679–702.
- . 2001a. Empirical regularities from interacting long and short memory investors in an agent-based stock market. *IEEE Transactions on Evolutionary Computation* 5: 442–455.
- . 2001b. Evolution and time horizons in an agent-based stock market. *Macroeconomic Dynamics* 5 (2): 225–254.

- LeBaron, Blake. 2001c. *A builder's guide to agent-based financial markets*. Technical report, International Business School, Brandeis University, Waltham.
- LeBaron, B. 2006. Agent-based computational finance. In *Handbook of computational economics volume 2: Agent-based computational economics*, Handbooks in economics series, ed. K.L. Judd and L. Tesfatsion. Amsterdam: Elsevier North-Holland.
- Parunak, H., Savit, R., and Riolo, R.L. 1998. Agent-based modeling vs. equation-based modeling: A case study and users' guide. In *Multi-agent systems and agent-based simulation*, ed. MABS, C. Sichman, and Gilbert. Berlin: Springer.
- Rescher, N. 1980. *Induction: An essay on the justification of inductive reasoning*. Pittsburgh: University of Pittsburgh Press.
- Tay, N.S.P., and S.C. Linn. 2001. Fuzzy inductive reasoning, expectation formation and the behavior of security prices. *Journal of Economic Dynamics and Control* 25 (3, 4): 321–361.
- Tesfatsion, L. 2002. Agent-based computational economics: Growing economies from the bottom up. *Artificial Life* 8 (1): 55–82.



Khairy Tourk and Hui Li

Abstract

One major factor that led to the 1997 Southeast Asian financial crisis was the reliance of the afflicted nations on heavy borrowing from western banks. The crisis has shown the massive need for establishing a regional bond market. Given the huge foreign reserves held by Asian central banks, at present, it is crucial to create a vehicle in order to preserve Asian capital within the region. Recent progress has been made in the direction of creating regional bond markets in the areas of Asian Bond Fund (ABF) that deals in foreign currency and Asian Basket Currency (ABC) bonds that deal in local currency. It's been almost a quarter of a century since *the 1997 Asian financial crisis*. In the years since then, there have been a number of seismic shifts in Asia's financial markets. This has been reflected in major improvements in the issuance of Asian government bonds. Yet, the area of corporate bonds in the region still remains somewhat underdeveloped due to the lack of credit ratings at investment grade. Addressing the issue of ratings is one of the real challenges that must be overcome before the Asian region could have a viable bond market. Following the central banks' policies of quantitative easing that were deemed necessary to deal with the economic downturn of the Great Recession of from 2007–2009 and the covid-19 pandemic, capital flows to the region had grown rapidly. Asian policy makers, however, should not lower their guard because the flows can be reversed at any point by external conditions.

Keywords

Executive Meeting of East Asia and Pacific (EMEAP) central banks · Takatashi Ito · Thansim Shinawatra · First Asian Bond Fund (ABF-1) · ABC bond corporation · The Asian Bond Market Initiative (ABMI) · Credit enhancement ·

K. Tourk (✉) · H. Li
Illinois Institute of Technology, Chicago, IL, USA
e-mail: tourk@stuart.iit.edu; hli130@hawk.iit.edu

Collateralized Bond Obligation (CBO) · Securitized asian corporate bonds · International rating agencies · Belt and road initiative · AIIB · Foreign investors · Emerging markets and rich nations' monetary policies · SOEs · Financial regulators · Yuan internationalization

42.1 Introduction

Before the Southeast Asia financial crisis, the standard works on the role of capital markets in economic growth (Goldsmith 1965, 1985; Hakansson 1992; McKinnon 1971, 1991) had focused mainly on the equity markets. Thus, the establishment of bond market in emerging nations has not received the attention it deserves in the finance literature.¹ For decades, the emerging economies in Asia have grown rapidly without the presence of an active bond market (Dalla et al. 1995; Emery 1997; Levine 1997; Sheng 1994; Yam 1997). In these economies, bond markets are very small relative to equity markets and the banking system.² One of the reasons that enticed Asian firms to borrow heavily from western banks before 1997 was the lack of an Asian bond market.³ A major consequence of the crisis was to breathe life into the concept of setting up an Asian bond market in order to enhance financial stability (Alba 1999; Rhee 2003; Tourk 2004).

A viable bond market gives the society a true measure of the opportunity cost of funds. In the absence of such a market, there is a loss of an important signal to channel savings into proper investments.⁴ Another drawback is the distortion of the incentives for risk taking, which raises the specter of a banking crisis.⁵

¹ Boot and Thakor (1997), pp. 693–733.

² Herring and Chatusripitak (2000), p. 1.

³ At the end of 2001, the total outstanding amount issued by Asian emerging borrowers in local and international capital markets showed rapid increase, reaching \$1.1 trillion. According to the IMF, in comparison with other regions, “emerging Asia has issued less international bonds (13 percent of the Asian total outstanding), more government bonds in local markets, but not so dominating (50 percent), and a large private sector share: 20 percent for financial institutions and 17 percent for corporate. The composition of the Asian bond markets is thus similar to that of mature bond markets: domestic governments bonds account for 49 percent, financial institutions 22 percent, corporate 11 percent, and international bonds 18 percent. By contrast, Latin America has issued more international bonds (32 percent – mainly by sovereign borrowers) and government bonds in local markets (55 percent), leaving a small share for the private sector (13 percent). Similarly, EMEA (Eastern Europe, Middle East, and Africa) has also issued more international bonds (23 percent) and government bonds in local markets (73 percent), with a negligible share of 4 percent for the private sector.” Tran and Roldos (2003), p. 2.

⁴ For example, lacking such signal, firms would overinvest if their internal rate of return is too low, as happened in many emerging economies in Asia during the 1990s.

⁵ This is particularly true when banks are undercapitalized. In such cases, making a write down in a loan renegotiation might result in violating capital adequacy standards. To avoid declaration of default, some banks might find it expedient to continue funding borrowers with negative present value projects. This exposes the banking system to liquidity shock. Without a bond market, banks do not have the option of selling bonds in a secondary market. Thus, they are likely to accept large losses on the sale of bad loans.

Overdependence on bank borrowing means that banks can extract rents from the borrowers as the cost of borrowing is higher than the case where firms have the option of accessing funds from a bond market.⁶ Long-term investment is also biased under bank-centered loans since banking institutions tend to offer loans for periods that are significantly shorter than the life span of long-term bonds.

The concept of an Asian bond market is not a new one. The original idea was proposed by the Asian Development Bank, which in the early 1990s issued dollar-denominated “Dragon Bonds.”⁷ The target of these bonds was the Asian investor. Unfortunately, the bonds did not succeed in attracting enough demand because of their low liquidity.

After 1997, the idea of an Asian bond as a vehicle to preserve long-term Asian capital within the region was a direct response to the financial crisis. The new initiative was taken by Japan’s Ministry of Finance (MOF). It was supported by the United States. Thanks to the enthusiastic promotion by the Thai Prime Minister, Thaksin Shinawatra, the Asian bond market has become a reality. He floated the idea in October 2002, at the East Asian Economic Summit organized by the World Economic Forum. Stemming from this, an Asian Bond Fund was established in June 2003, by the heads of central banks in the region.

The *raison d’être* of the fund is for Asian governments and other entities to issue bonds in order to reinvest part of the region’s savings in Asia itself (Oh 2003; Sonakul 2000). “Since bond rating and settlement is to be handled within the region, the use of dollar bonds will become unnecessary. Eventually, it should be possible to issue bonds denominated in regional currency or a basket of regional currencies.”⁸

Developing local bond market (Mungthin 2000) is important in reducing the Asian country’s exposure to maturity and exchange rate risks, and “sudden stops” in the availability of international capital.⁹ The benefits include an increase in the efficiency of allocating surplus funds and the retention of Asian capital. In 2004, central banks of Asia held roughly \$1.1 trillion of US Treasuries. In just two years, beginning at the start of 2002, the dollar dropped around 26% against a basket of six major currencies.¹⁰

For Asian corporations, they benefit from the diversification of funding sources (i.e., less reliance on borrowing from domestic banks) and improving transparency, leading to better corporate governance. Another advantage is that it would give Asian governments more policy instruments to stabilize their financial markets.¹¹

⁶Richard J. Herring and Nathporn Chatusripitak, op. cit. pp. 2125.

⁷These dragon bonds were simultaneously issued in more than two Asian markets. The markets included the “dragon” economies of Singapore and Hong Kong. Nagao (2003), p. 39.

⁸Makoto (2003), p. 34.

⁹Tran and Roldos (2003), p. 1.

¹⁰Pesek Jr (2005), p. B4.

¹¹Nils H. Hakansson, “The Role of Corporate Bond Market in an Economy – and in avoiding crisis”, *Working Paper*, University of California, Berkeley.

The Asian Bond Market has taken two forms: first an Asian bond market led by East Asia-Pacific (EMEAP) Central Banks, and the second is represented by the Asian Bond Market Initiative (ABMI), proposed in 2002 by Japan's MOF, and under the supervision of the ASEAN +3 finance ministers.

42.2 The Asian Bond Market Launched by EMEAP Central Banks

At the beginning of 2005, the foreign exchange reserves in the Asian regions were around \$2 trillion dollars, which equals more than 50% of the world's reserves.¹² These Asian resources could be channeled into banking, as well as other types of finance,¹³ to enable Asian nations to create wealth in each other's economy. Thus, the necessity to establish an infrastructure for bond markets, both primary and secondary. Recent progress has been made in the direction of creating regional bond markets in the areas of Asian Bond Fund (ABF) that deal in foreign currency and Asian Basket Currency (ABC) bonds that deal in local currency.¹⁴

The first Asian bond fund (ABF-1) was launched in June 2003, by the central banks of the 11 countries who are members of EMEAP.¹⁵ It consists of \$1 billion worth of foreign reserves, which is being invested in a basket of dollar-denominated bonds issued in eight Asian economies¹⁶ by sovereign and quasi-sovereign entities in the EMEAP countries, except in Japan, Australia, and New Zealand. The Bank of International Settlements (BIS) will be managing the fund. It has been indicated that the 11-member group plan to consider extending ABF investments to bonds denominated in Asian currencies. In December 2004, a second fund was launched with \$2 billion in governments' capital to invest in corporate debt issued with local currencies.¹⁷

¹²At the end of March 2003, China, Hong Kong SAR, Japan, Korea, Singapore, Taipei, Thailand held \$662 billion worth of U.S. treasury securities. This represented 50% of total foreign holdings of these securities and 20% of total outstanding treasury securities. These reserves are highly excessive with respect to the seven East Asian nations import needs and exchange rate management. Kiang 2003, cited in Rhee 2003, p. 9.

¹³According to one analyst, the 1997 crisis would have completely different consequences had the Asian region allocated 15% of its accumulated reserves toward the capital account crisis. S. Ghon Rhee, *Second Annual Conference of the PECC Finance Forum*, op. cit., p. 6.

¹⁴This section draws heavily on S. Ghon Rhee, *Second Annual Conference of the PECC Finance Forum*, op. cit. pp. 9–12.

¹⁵The 11 EMEAP consists of China, Hong Kong, Indonesia, South Korea, Malaysia, Philippines, Singapore, Thailand, Japan, Australia, and New Zealand.

¹⁶The eight economies are those of China, Hong Kong, Indonesia, South Korea, Malaysia, Philippines, Singapore, and Thailand.

¹⁷Steven Glain, "Asia Creeping Unification", *Newsweek*, February 7, 2005, p. 32.

42.3 The Asian Bond Market Initiative (ABMI)

The Asian Bond Market Initiative (ABMI) is focused on the creation of a proper environment where access by a wider variety of issuers to the regional bond markets is facilitated (Taniguchi 2003). Since summer of 2003, working groups representing the ASEAN +3 finance ministers have defined specific subjects to be reported on to the ASEAN +3 finance ministers meeting in 2004.¹⁸

One important step in the creation of proper environment is for the government to establish a benchmark yield curve to serve as the risk-free rate for the pricing of other securities.¹⁹ A program of regular issues at appropriate maturities (e.g., three month, six month, one year, three years, five years, and eventually ten years) should be set up to 20. The interest rate on government bonds should be market determined. It must be kept in mind that minimizing the cost of government borrowing may be in conflict with the development of this market-determined rate. Thus, the government should refrain from any manipulation of the bond market (i.e., requiring some institutions to hold government debt, devising preferential tax treatment for public sector debt) with the purpose of reducing the cost of its borrowing. Doing that would negatively affect the efficiency of the bond market and lead to distortions in the allocation of capital.²⁰

42.4 The ABC Bond Corporation

The creation of the ABC bond corporation, dealing in local currency, is the result of a Japanese MOF initiative. One major benefit of the ABC government bonds is that they represent a benchmark for the region. Ito (2003a, b) proposed the establishment of this corporation in order to serve as a depository for financial assets supplied by participating governments in the form of local currency-denominated government bonds.

For the ABC government bonds to be priced correctly, it is important that the public sector should be vigorously involved in establishing a deep domestic bond market. Governments must cooperate to issue standardized bonds that can be put into depository to form a standardized basket (i.e., standardized maturity; standardized interest rate calculation, preannounced coordinated issuance schedule, interest payment methods, and depository location).²¹ To reduce the weight of the dollar in Asian foreign reserves, ABC bonds that are issued offshore will be treated as foreign reserves.

¹⁸Hisatsugu Nagao, op. cit. p. 39.

¹⁹In practice, government issues provide the benchmark for estimating term structure of interest rate. These issues are approximately default risk-free in domestic currency terms.

²⁰Richard J. Herring and Nathporn Chatusripitak, op. cit. p. 25.

²¹Takatoshi Ito, Asian Basket Currency(ABC) Bonds, February 2003.

In the first phase, the ABC Bond Corporation would issue bonds that match the value of the assets. In the second phase, the corporation may issue bonds that match the value of corporate bonds denominated in various Asian currencies. As such, the ABC Bond Corporation operates along the same lines as a Special Purpose Vehicle (SPV) with asset-backed securities. It is here that the private sector can establish a corporate bond market and asset-backed securities denominated in the basket currency. Corporate bonds are brought into the depository, just like government bonds, and ABC corporate bonds can be issued. The benchmark yield is provided by the ABC sovereign bonds. Because the credit risk is higher, the yield would also be higher. Furthermore, as these bonds can now be sold in the region at large, instead of in one country alone, this would reduce the issuing costs. Another advantage is the direct relationship between investors and borrowers.

One major attraction of issuing ABC bonds is that they will diversify currency risk. Firms exporting in the Asian region will issue liabilities that match their revenue streams denominated in the local currencies. By being less dependent on bank loans, they will be less affected by banking crisis.²²

The ABC Bond Corporation operates as an issuer of bonds to be bought by Asian institutional investors. Expected buyers of the ABC bonds include Asian central banks and pension funds. They also include institutional investors, both Asian and foreign. This is different from the ABF, which serves as an investment vehicle for Asian central banks, acting as a buyer of dollar-denominated bonds issued by Asian borrowers.

It should be kept in mind that, since pension funds and central banks tend to have passive investment strategies, both the ABF and ABC invested bonds might not be powerful enough instruments in the promotion of market liquidity on the secondary market.²³

42.5 Credit Enhancement

As explained above, the bonds issued by the ABC Bond Corporation match the value of the underlying assets of this entity, which are denominated in government bonds of various Asian currencies. Even though pledged bonds are backed by full faith and credit of participating Asian governments, the question of credit enhancement comes into play because international rating agencies view various governments differently with respect to credit risk. The enhancement could be either *internal* or *external*.

The most used form of internal enhancement is through the restructuring of the distribution of pooled cash flows to create a new instrument called a Collateralized

²²Takatoshi Ito, "Promoting Asian Basket Currency (ABC) Bonds," RCAST, University of Tokyo, March 1, 2003.

²³For the trade-off between liquidity and foreign exchange risk in the case of basket-denominated bonds, see Ogawa and Shimizu 2004, pp. 719–738.

Bond Obligation (CBO). Under this system, two tranches of ABC bonds could be issued: senior bonds with higher credit rating and lower yields and subordinate bonds with lower credit rating but higher yields.²⁴

External enhancement includes third-party guarantees provided by official financial institutions (e.g., Development Bank of Singapore, ADB, and JBIC), letters of credit by leading international commercial banks, and bond insurance provided by monoline bond companies.²⁵

In order to transform Asian corporate bonds from speculative-grade assets to investment-grade assets, Asian governments, Asian banks, and Asian credit insurers could offer full or partial guarantees on the issuer. This will facilitate the issuance of credit insurance by top-rated insurers to bond issuers with poor credit ratings. Asian governments could also play an important role in strengthening a system of Asian-based credit insurers.²⁶ China has made important progress in this regard. For example, its domestic rating agency, China Lianhe Credit Rating Company had won an award as the best domestic rating agency in China.²⁷

42.6 Securitized Asian Corporate Bonds

The past few years have seen major improvements in the issuance of Asian government bonds. Yet, the area of corporate bonds in the region still remains clearly underdeveloped (Claessens et al. 1998) due to the lack of credit ratings at investment grade.

One way to raise the quality of Asian corporate bonds to investment-grade assets is to pool the bonds together making them securitized corporate bonds. A large asset pool of this kind can be listed and traded on the exchange or in the OTC market. This has the advantage of satisfying the market demand for high-yield assets with credit protection. These assets are considered safe because top-rated credit insurers ensure the cash flow from the bonds.

²⁴This type of senior/subordinated bond issue was extremely popular in the post crisis period in Korea (see Ohet al. 2003).

²⁵Many monoline bond insurance companies operate in the United States and Europe providing municipal bond insurance and pool insurance.

²⁶The first regional bond insurance company was established by the Asian Development Bank (ADB) in Singapore in 1995. This company was named as the Asian Securitization and Infrastructure Assurance Ltd. (ASIA Ltd). Its owners consist of a diverse group consisting of financial institutions, ADB member countries, insurance companies in addition to others. Primarily, ASIA Ltd. offers credit insurance to infrastructure projects of developing nations. Richard Yan Ki Ho and Chak Sham Michael Wong, "Road Map for Building the Institutional Foundation for Regional Bond Market in East Asia", *Second Annual Conference of the PECC Finance Forum*, op. cit. p. 21.

²⁷The award was given by FinanceAsia, a financial markets publication based in Hong Kong that is widely followed by Asia-Pacific market participants. The awards were judged by its editors and an advisory board of prominent investors and bankers. "Winners: FinanceAsia's Country Awards 2021, Part 1", May 25, 2021.

So far, Asian securitization has developed very slowly in the region.²⁸ At present, the investor base for Asian corporate bonds is extremely narrow. While credit enhancement would increase the participation of both Asian institutional and international investors, there is also a need to attract *individual* investors to be active participants in Asian debt. In this regard, the government policies could be important in developing bonds of smaller par value that the general public is enticed to invest in.

Another policy is government support of mutual funds that invest primarily in Asian corporate bonds. An encouraging development is the establishment of the Asian bond fund, whose objective is to invest in Asian debt securities. It was established in May 2003, by nine Asian governments, where each contributed up to one percent of its foreign reserve to the fund. These governments consist of Thailand, China, Japan, Hong Kong SAR, South Korea, Singapore, Indonesia, Malaysia, and the Philippines. It has been suggested that in order to encourage the participation of individual investors,²⁹ the government bond fund can be further securitized to become either a closed or open-ended fund.³⁰ Assets in the fund must be insured in order to minimize the credit risk of individual investors. Due to its large market capitalization and prospective large number of investors, both institutional and individual, the liquid government bond fund can flourish into an actively traded asset capable of nurturing a number of financial derivatives. Such derivatives would enable international investors and credit insurers to hedge the credit and market risk in Asia. In short, the Asian governments can play a crucial role in asset securitization that would transform sovereign and corporate debts into attractive investment vehicles for the general public.³¹

²⁸ Korea has been an exception, being relatively active in this field. Korea set up a system of Primary Collateralized Bond Obligation in the year 2000. This program was introduced via the Korea Credit Guarantee Fund. In Hong Kong, the Mortgage Corporation, a subsidiary of the Hong Kong Monetary Authority, deals with mortgage securitization. Korea has enacted the Securitization Act in order to provide a workable legal system to facilitate the securitization of banks' nonperforming loans in addition to restructuring the balance sheet of financially troubled corporations.

²⁹ For the benefit of institutional investors, the Hong Kong Monetary Authority issues Exchange Fund Bills and Notes. They are akin to U.S. treasury securities. The Authority started to issue smaller Exchange Fund Notes of around \$7000 per contract in 2000. These notes are popular among the general public.

³⁰ Richard Yan Ki Ho and Chak Sham Michael Wong, "Road Map for Building the Institutional Foundation for Regional Bond Market in East Asia", *Second Annual Conference of the PECC Finance Forum*, op. cit. p. 26.

³¹ During the South East Asian Crisis, the Hong Kong government took a pro-active stand against the attacks of global hedge funds speculators. To protect the stock market, the "equities purchased at that time were securitized to become a fund known as Tracker Fund. All the units of the Fund were later sold to both institutional and individual investors." Richard Yan Ki Ho and Chak ShamMichaelWong, "Road Map for Building the Institutional Foundation for Regional Bond Market in East Asia", *Second Annual Conference of the PECC Finance Forum*, op. cit., p. 26.

42.7 Efficient Financial Intermediaries

An integral part of creating bond markets is the development of an efficient pension and insurance system. The main advantage of the financial intermediaries, such as pension funds and insurance companies, is the ability to invest in financial instruments other than government bonds. Creditworthy Asian corporations would be able to borrow long-term. In addition to expanding the finance choices for the private sector, many governments in Asia might find it preferable to develop a yield curve in order to meet their budget deficit obligations. Except for Japan, which has a well-developed bond market, there are still many barriers that must be overcome before it becomes a reality in other parts of Asia.³² The barriers include weak financial institutions, restrictive investment eligibility requirements, and antiquated transfer, trust, title, and tax policies.³³ Another major difficulty is the reluctance of international rating agencies³⁴ to analyze the creditworthiness of Asian corporations. Many of these corporations are family owned and there is a lack of information regarding their financial standing. Thus, poor credit ratings, as assessed by the dominant international rating agencies, are one of the main reasons preventing Asian corporations from issuing unsecured bonds.³⁵

42.8 Is the Classic Form of Emerging Market Debt Crisis a Thing of the Past?

Asian economies have come a long way since the Asian financial crisis in 1997 and the sub-mortgage crisis in 2008. The next analysis points to a definite improvement in present-day financial picture. Asia has not experienced another serious crisis for more than a decade. However, our analysis will show that the specter of a future currency crisis is still present. Thus, Asian nations should not be lulled into complacency.

³²In 2003, Japan had a \$5 trillion market, which is roughly equal to its gross domestic product. China, on the other hand, has less than \$450 billion in issues outstanding representing 32% of its GDP. Mays and Preiss 2003, p. 15.

³³Marshall Mays and Michael Preiss, op. cit., p. 15.

³⁴In the long term there is a need to establish Asian rating agencies. Practically however, it will take a few decades before the dominance of international rating agencies could be seriously challenged.

³⁵It has been estimated that for Asian corporations with less favorable credit ratings, the funding costs of issuing unsecured Asian corporation bonds could be higher than 18%. This is why many corporations in Asia prefer to acquire collateral pledged bank loans at an interest rate varying between 5 and 10%. Richard Yan Ki Ho and Chak Sham, Michael Wong, "Road Map for Building the Institutional Foundation for Regional Bond Market in East Asia", *Second Annual Conference of the PECC Finance Forum*, op. cit. p.16.

42.9 Recent Trends in Global Finance

42.9.1 Ample Liquidity

Before analyzing the recent trends in the Asian bond market, it is important to emphasize the fact that both the west and Asia are awash in liquidity. In the USA, money market funds in 2020 looked after \$4 trillion of savings for individual and businesses.³⁶ In addition, the S&P 500 nonfinancial firms have cash holdings that are around \$1.4 trillion in 2021.³⁷



The increase in liquidity in the west is due in part to the expansionary monetary policy enacted after the 2009 financial crisis and the 2020 pandemic. The situation in Asia, however, is different. The high level of its liquidity is simply a result of high rate of savings among Asian households. China is a case in point. In 2020, the rate of saving was estimated at \$18.8 trillion and is expected to rise to \$51 trillion in 2030.³⁸

The vast pool of liquidity resulted in short-term US Treasury bills having negative interest rates in 2020. In China, due to capital controls, the interest rate that savers received domestically is pitifully low. Such a situation points to the need for a healthy bond market that, among other things, meets the needs of cash holders for positive returns.

³⁶Smith and Rennison (2021).

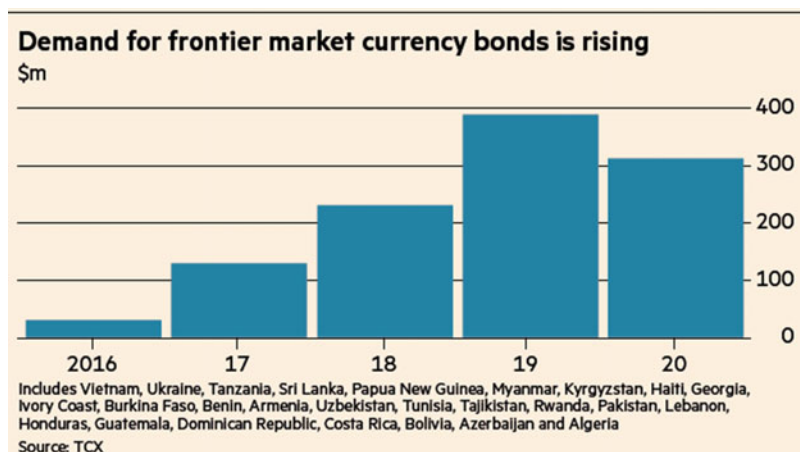
³⁷“The coming capex carnival”, *the Economist* 2021b.

³⁸“Chasing a pot of gold”, *the Economist* 2021a, p. 63.

42.9.2 The Demand for Emerging Market Currency Bonds Is Rising

According to the ADB report, during the past decade, Asia has attracted global attention as a growth engine and successfully become an indispensable power in the world economy.³⁹ In 2020, the combined GDP of China, Japan, and Korea was around \$21 trillion, accounts for 25% of the world economy and their foreign reserves exceed \$4.97 trillion. Moreover, ASEAN countries GDP are around \$3 trillion. Having the most dynamic economy in the world means that the Asian bond market is going to be a major pillar of the global finance industry.

The following diagram shows the strong appetite of investors for investment opportunities in emerging markets.⁴⁰



42.9.3 Asset Management

Even though at present, asset management in Asia is still young, yet, it is only a matter of time before home-grown asset managers would be in a position to acquire the skills needed to generate relatively respectable returns. Here, the role of foreign asset managers is expected to grow in Asia.

Just as East Asian has been able to master industrial technologies in a relatively short time, it is likely that Asian asset managers could do the same in coming decades. China is a case in point. The country is expected to have \$70 trillion of

³⁹“Asian bond monitor March 2021”, *Asian Development Bank*.

⁴⁰Szalay (2021).

household investable assets by 2030.⁴¹ The following diagram shows the growth in total assets under management.



42.10 Asian Finances in the 1997 and 2021: A Comparison

While developing a deep and liquid bond market in Asia that matches that of the USA is still a far-off proposition, there are positive developments that bode well for the formation of a bond market that is promised to be more stable and mature than the one Asia had at the end of the twentieth century. Many of these developments stand in stark contrast to the situation prevailing around the 1997 crisis. In this section, we will provide a comparative analysis that juxtaposes the financial situation in Asia in 1997 with the present.

42.10.1 From Capital Importer to Capital Exporter

In 1990s, except for Japan, Asia was still a capital importer. This situation has drastically changed. At present, Asia has a surplus capital. Two factors are responsible for the change. Since 2010, China has become a capital-exporter. Secondly, the 1997 crisis made Asian Central Banks cognizant of the importance of increasing their foreign reserves.

⁴¹ Pesek (2021).

42.10.2 Short-Term and Long-Term Capital Flows

The 1997 crisis was caused by speculative activities in Western financial centers. In the early 1990s, many Asian nations have adopted the “Washington Consensus” to open their cross-border capital markets. At the beginning, such policies seemed to work well as many of them, such as Thailand, Malaysia, and Indonesia, welcomed foreign investments, which contributed to their rapid economic growth. However, such heavy inflows of foreign investments made these countries vulnerable to foreign investors as the countries were heavily in debt and their continued growth depended largely on the willingness of foreign investors to provide funding. This resulted in a financial crisis that bubbled under the surface. More specifically, one reason behind the crisis was that many Asian firms borrowed short-term but invested in long-term. The second reason was that even though most of the afflicted countries had sound economic fundamentals, the paucity of their foreign reserves made it easier for speculators to short their currencies.

Nowadays, Asian Central Banks have large foreign reserves. While holding large amounts of reserves could help weather currency-related crises in the emerging economies, it also comes with a cost. Most of the foreign reserves are invested in US Treasuries, which only earn a modest return. In the meanwhile, the extensive use of unlimited quantitative easing monetary policy since the 2009 financial crisis has kept diluting the total value of these countries’ foreign reserves.

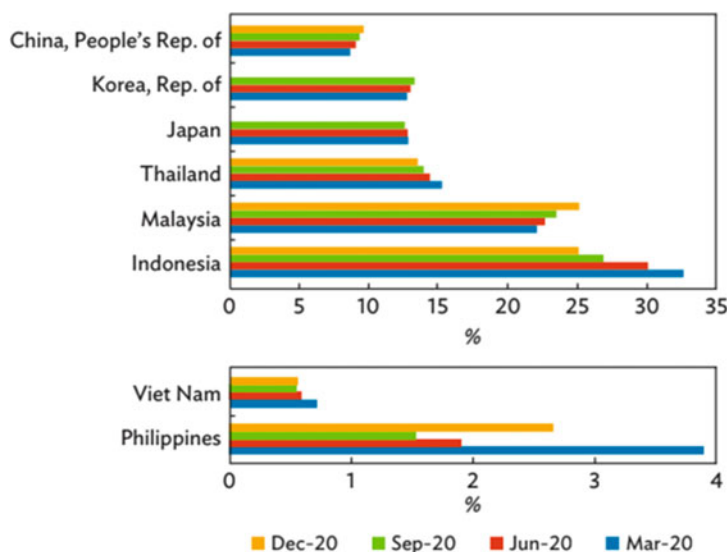
42.10.3 Different Investor Motivations

Before 1997, short-term capital movements motivated by the prospects of making profits on currency speculation dominated foreign investors’ decisions. Nowadays, Asia is attracting more long-term investors.

Rather than being solely regarded as global production bases, Asian countries are now heading toward to develop a global market with strong purchasing power. Swarms of multinational companies are now investing in Asian manufacturing facilities, produce goods, and then sell them in the local market.

The strength of their economies has made *sovereign bonds* attractive to foreign investors. The following diagram shows the percentage of foreign holdings in *local* currency government bonds in the Asian market, during the period of March to December 2020.⁴²

⁴²“Asian bond monitor March 2021”, *Asian Development Bank*, p. 5.



Note: Data for Japan and the Republic of Korea as of 30 September 2020.
Source: AsianBondsOnline.

For example, in China, by March 2021, yuan-denominated bonds held by foreigners were \$551 billion (3.55 trillion yuan). They have grown nearly 60 percent over the year 2020.⁴³ The reason for this exceptional growth is that the interest rate in China on 10-year bonds was higher than that in the west. Most of the demand is on the part of private institutional investors, central banks, and others (e.g., Japan Government Pension Investment Fund).

42.10.4 Asian Regulators and Financial Market Stability

Since 1997, Asia has not suffered any major financial crisis. This is due in part to the rise in the quality of financial regulators. Many Asian countries have banned or limited offshore transactions of their currencies and offshore issuance of bonds denominated in their currencies amidst growing concerns that those transactions might destabilize the currency's value.

In China, financial regulators, through *selective targeting* policies, have proved over and over again their ability to nip most bubbles in the bud. For example, in 2015, the increase in demand for Chinese assets by foreigners caused the yuan to strengthen, with negative implication for exports. To help cool the yuan rally, Beijing had approved a record amount of investments to flow out of the country, which allows investors to access outside mainland China through banks and other

⁴³ Kawate (2021).

institutions. Recently, in June 2021, a cumulative \$147 billion of approval has been added to the nation's qualified domestic institutional investor scheme, which allows investors to access assets *outside* mainland China through banks and other institutions.⁴⁴

Another example is that in June 2020, Beijing came up with a scheme that is set to be a milestone in the liberalization of China's capital control. The Wealth Connect program would allow up to \$23 billion (150 billion yuan) to be invested in either direction between Hong Kong and mainland China, with limit for individuals of \$155,000 (one million yuan). This pilot program covers 70 million people living in the cities of Guangzhou and Shenzhen.⁴⁵

42.10.5 The Increasing Importance of Asian Financial Hubs

Unlike the 1990s, the center of economic gravity has now moved from the west to the east. Yet, in the area of finance, Asia does not have sophisticated financial institutions and suffers from a lack of financial expertise. On the other hand, the growth in Asian economies is increasing the importance of Asian financial hubs. Other than Hong Kong and Singapore, new financial centers are rising in cities like Shanghai and Shenzhen.

42.10.6 Geopolitics

The role of Asia in global finance is growing rapidly. One major feature of such growth is the increasing links between the USA and China. This is happening in an environment where geopolitical tension is rising between the two largest economies in the world. Before 2010, the Washington-Beijing relationship was relatively friendly. This is not the case now with the two nations' relationship changing from partnership to competitors. The question that poses itself is that whether the rising tension between the two largest economies would spill over into economic matters. Judging from the limited evidence so far, it seems that decoupling is not the wave of the future. Except in a few areas related to security, all indications in terms of foreign direct investment and the capital flows point to the coupling trend not being seriously affected by the Cold War rhetoric. With the titans of American finance being deeply embedded than ever into China, this bodes well not only for the two largest economies in the world but also for world peace.

⁴⁴Hale and Lockett (2021a).

⁴⁵Hale and Kinder (2021).

42.11 Long-Term Trends

42.11.1 The Increasing Role of Large Projects

The demand for long-term bonds is expected to rise significantly in the coming decades on the part of emerging economies in Asia. One of the major reasons is that Asian industrialization would require huge funding to finance new special economic zones and other capital-intensive projects (e.g., infrastructure). This is particularly true for the members of China's new Silk Road. Another important reason is the demand for green bonds, which is expected to raise substantially both in developing and developed nations.

42.11.2 Yuan Internationalization

Yuan internationalization has the advantage of reducing currency risk. Given the growing intra-trade in Asia and with China being the largest trading partner for most Asian economies, it is expected that many of the trade transactions and investment projects would be conducted in the yuan. This is particularly true in the case of China's Belt and Road Initiative, where financial integration is part of the country's mega project. Thus, taking the long view, the internationalization of the yuan looms large on the horizon as a dollar alternative for many Asian economies.

42.11.3 Financial Technology Transfer

Foreign bond holders are expected to play an important role in the evolving financial market in Asia. Many Asian economies are experiencing demographic changes, as reflected in a rising portion of the elderly in the total population. This results in an increasing demand for competent asset managers whose skills are needed to provide respectable returns that would ease the financial burden of the expected increase in the cost of social programs. Foreign asset managers bring both capital and expertise.

One major recent development that bodes well for a brighter future is the recent liberalization in China's capital market. Thanks to China's capital account reform, the country has embarked on a policy to gain the necessary skills for asset management. For example, in April 2021, Goldman Sachs and BlackRock have received approval for a wealth management partnership with China Construction Bank. In June of the same year, Goldman Sachs had a wealth partnership with the Industrial and Commercial Bank of China. JPMorgan Asset Management in the meanwhile announced a plan to invest \$4.1 billion in China Merchant Bank's wealth unit. The European outfit Amundi and Schroders have also gained approval for majority-owned partnership in wealth management.

These developments allow the giants of western finance to draw on the huge savings of Chinese customers. Undoubtedly, their expertise would help the country to build efficient institutions capable of achieving positive returns that are not only enjoyed by the present generation but also would help to secure better living standards for the aging Chinese population.

42.11.4 State Financial Backing Is No Longer Assured for Chinese SOEs

Due to interest rate differential, Chinese investors have accumulated dollar debt on the international market. More than hundred billion dollars are coming due in 2021.⁴⁶ Global investors are a major source of dollar financing for Chinese firms.

In 2020, Chinese companies defaulted on \$7.3 billion of dollar-denominated offshore bonds. Investors in offshore bonds include Peking University Founder Group, a state-back conglomerate, and the Huarong Asset Management, which is owned by China's finance ministry. Unlike the past, the Chinese authorities do not now stand behind indebted public firms. The prospect of default has forced international lenders to reassess how much Chinese government support, if any, they will receive.

An example of the strict new policy is that of Huarong Asset Management, China's largest distressed debt manager. In 2020, it owed \$22 billion in dollar-denominated debt. The next diagram shows how Huarong bonds plunged on repayment fears with its dollar bond trading at 66 cents.⁴⁷



⁴⁶Hale and Lockett (2021b).

⁴⁷The company's former chair was executed in January 2021 for corruption. Hale and Lockett (2021b).

42.12 Conclusion

Compared to the situation in 1997 financial crisis, Asian nation economies now are in relatively better position to deal with future crises. This is due to the growth of foreign reserves and the improved quality of financial regulators. However, one should keep in mind that, due to the lack of well-developed financial markets, Asian nations are still subject to headwinds from the rich ones.

Specifically, Asian economies are not immune to the destabilizing effects of American monetary policy. A case in point is that, in 2013, the head of the Federal Reserve, Ben Bernanke, announced that the Fed *might* start “tapering” its quantitative easing policy. The effect of this casual remark resulted in global markets going into a tailspin which was called “Taper tantrum.” This took the form of increased interest rates, tumbling of risky assets and emerging economy *currency crashing*.

In 2021, the recent increase in supply of money in rich nations, such as the USA has resulted in low returns. This has made emerging markets attractive for investors looking for high yields. The real worry of emerging markets is that capital flows could *revert* abruptly, leaving their assets and currency in *turmoil*. The fact is that even though Asia owns huge foreign reserves, which could be used as a first line of defense against volatile capital flows, Asian nations still have to face *the impossible trinity*, which states that no economy can simultaneously have a fixed exchange rate, an open capital account, and an independent monetary policy. As such, there is really no protection against the destabilizing harmful effects in emerging country assets and currency markets.⁴⁸ In other words, Asian nations could still suffer from a possible shortage of dollars, which would inhibit their ability to repay foreign debts.

One major concern is that despite the borrowing costs going down substantially for rich nations as a result of loose money supply, the borrowing costs for emerging economies have been much higher. These nations spend much more in debt than advanced nations. For example, in 2019, advanced economies have borrowed roughly three times as much as emerging economies, yet they ended up paying as much interest as the emerging economies.⁴⁹

The following diagram shows the share of government revenues spent on foreign public debt interest payments.⁵⁰

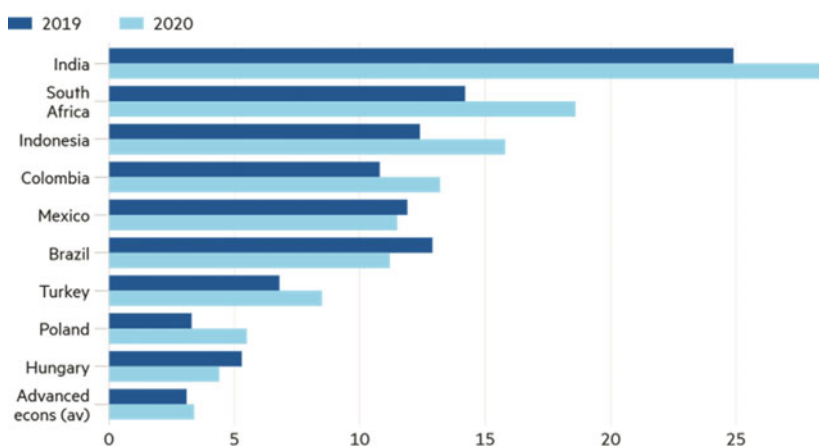
⁴⁸ Brazil offers a cautionary tale. Even though it has one of the largest foreign reserves in the world in 2014, it was forced to depreciate its currency in the face of foreign capital flows reversals.

⁴⁹ Wheatley (2021).

⁵⁰ Wheatley (2021).

Middle income countries spend more on debt than rich countries

Share of government revenues spent on public debt interest payments, %

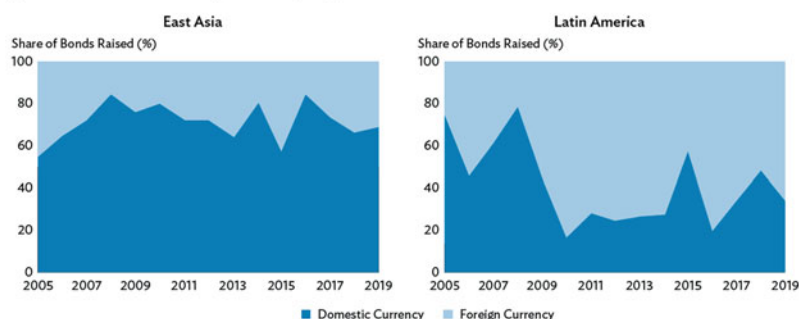


Source: Fitch Ratings
© FT

On the other hand, it should be pointed out that Asian economies are in a relatively better position than other developing nations. This is because the average rate of interest on Asian debts tends to be lower than their growth rates. In general, emerging economies are well advised to minimize short-term debt in foreign currency, especially sovereign debt.

Another advantage for Asia is that the share of domestic corporate bonds as the percentage of total corporate bond issuance is higher than other emerging markets. The following diagrams compare the situation in East Asia with that of Latin America.⁵¹

Figure B1: Domestic and Foreign Currency Corporate Bond Issuance



Note: This figure shows, for each region, the share of corporate bond financing raised in local and foreign currency for the median economy each year.
Source: Refinitiv's SDC.

⁵¹“Asian bond monitor March 2021”, *Asian Development Bank*, p. 13.

This is because developing local bond markets is an important step to reduce the dependence on the dollar. The ability of middle-income countries to sell debt in their own currencies would help to protect them from the damage that afflicted them in past crises. Empirically, however, the resilience of the local debt market is yet to be tested.⁵²

Acknowledgments The authors are indebted to Junji Nakagawa, Yoshinori Shimizu, and Akira Suehiro for valuable insights. We are grateful for stimulating discussions with Omar Farooq, Qi Liang, Ryusuke Shimizu, Kumiko Sakaim, Nariatsu Tanaka, Weiham Ni, Shiqi Li, Kun Li, Shouhao Li, Arthur Lin, and Michael Wang. The usual *caveat* applies.

References

- Alba, P., L. Hernandez, and D. Klingebiel 1999. *Financial liberalization and the capital account: Thailand 1988–1997*. Working Paper, World Bank.
- Asian Development Bank. 2021. *Asian bond monitor March 2021*, March 2021.
- Boot, A., and A. Thakor. 1997. Financial system architecture. *Review of Financial Studies* 10 (3): 693–733.
- Claessens, S., S. Djankov, J.P.H. Fan, and L.H.P. Lang 1998. *Corporate diversification in East Asia: The role of ultimate ownership and group affiliation*. Working Paper, World Bank.
- Dalla, I., D. Khatdhat, D.C.K. Rao, Kali, L. Jun, and T. Chuppe. 1995. *The emerging Asian bond market*. Washington, DC: The World Bank.
- Economist. 2021a. *Chasing a pot of gold*, June 12, 2021, p. 63.
- . 2021b. *The coming capex carnival*, May 29, 2021.
- Emery, R.F. 1997. *The bond markets of developing*. East Asia, Boulder, Colorado: Westview Press.
- Goldsmith, R. 1965. Chapter 2. In *The flow of capital funds in the postwar economy*. New York: National Bureau of Economic Research.
- . 1985. *Comparative National Balance Sheets: A study of twenty countries, 1688–1978*. Chicago: University of Chicago Press.
- Hakansson, N.H. 1992. Welfare economics of financial markets. In *The new Palgrave dictionary of money and finance*, ed. J. Eatwell, M. Milgate, and P. Newman, vol. 3, 790–796. London: Macmillan Press.
- Hale, Thomas, and Kinder, Tabby. 2021. China edges closer to wealth link with Hong Kong in boost for liberalisation, *Financial Times*, May 21, 2021.
- Hale, Thomas, and Lockett, Hudson. 2021a. China approves record outflows from local investors to cool renminbi rally. *Financial Times*, June 16, 2021.
- . 2021b. Chinese corporate dollar bonds hit as Beijing tightens credit. *Financial Times*, May 14, 2021.
- Herring, R.J., and N. Chatusritipak 2000. The case of the missing markets: the bond market and why it matters for financial development. *Asian Development Bank Institute*. Working Paper 11, Tokyo, July 2000, 23–25.
- Ito, T. 2003a. *Promoting Asian Basket Currency(ABC) bonds*. RCAST, University of Tokyo, 1 March 2003.
- . 2003b. *Asian Basket Currency (ABC) Bonds*. February 2003.

⁵²According to David Lubin, head of emerging market economics at Citi, “we don’t have any experience of a domestic debt crisis in a large emerging market. . .the kinds of things that act as triggers for external financing problems may also act as triggers for domestic financing problems.” Wheatley (2021).

- Kawate, Iori. 2021. China's capital inflows hit 7-year high on bond-buying bonanza, *Nikkei Asia*, May 21, 2021.
- Kiang, L.H. 2003. East Asian capital markets: challenging times, A keynote address delivered at the *Euromoney 10th Asia-Pacific Issuers & Investors Forum*, 18 March, Singapore.
- Levine, R. 1997. Financial development and economic growth: Views and agenda. *Journal of Economic Literature* 36: 688–726.
- Mays, M., and M. Preiss. 2003. Asia must put its savings to work. *Financial Times* 7: 15.
- McKinnon, R.I. 1971. *Money and Capital in Economic Development*. Washington, DC: The Brookings Institution.
- . 1991. *The order of economic liberalization: Financial control in the transition to a market economy*. Baltimore: Johns Hopkins University Press.
- Mungthin, N. 2000. *Thai bond market development (in Thai)*. Bank of Thailand.
- Nagao, Hisatsugu. 2003. Market for Asian bonds taking shape. *The Nikkei Weekly*, 24 November, 39.
- Ogawa, O., and J. Shimizu. 2004. Bond issuers' tradeoff for common currency basket denominated bonds in East Asia. *Journal of Asian Economics* 15: 719–738.
- Oh, G., D. Park, J. Park, and D.Y. Yang 2003. How to mobilize the asian savings within the region: securitization and credit enhancement for the development of East Asia's bond market. *Korea Institute for International Economic Policy*. Working Paper, 3 February 2003.
- Pesek, William. 2021. Wall street bulls charge into China's opening markets. *Asian Times*, June 7, 2021.
- Pesek, W. Jr. 2005. Dollar skeptics in Asia have prominent company. *International Herald Tribune*, 3 February 2005, B4.
- Rhee, S.G. 2003. The structure and characteristics of East Asian bond markets. *Second Annual Conference of the PECC Finance Forum*, on Issues and Challenges for Regional Financial Cooperation in the Asia-Pacific, Hua Hin, Thailand, 8–9 July, 21.
- Sheng, A. 1994. Future directions for Hong Kong's debt market. Speech at the *First Annual Pan-Asia Bonds Summit*, 29–30 November.
- Smith, Colby, and Rennison, Joe. 2021. Negative rates put \$4 trn savings industry under heavy strain, *Financial Times*, June 5, 2021.
- Sonakul, M.R.C.M. 2000. *Keynote address on the occasion of the ADB conference on government bond market and financial sector development in developing Asian Economies*. Manila, 28–30 March.
- Szalay, Eva. 2021. Hunger for yield pushes foreign investors into frontier territory, *Financial Times*, May 19, 2021.
- Taniguchi, T. 2003. Asian bond market initiative. A keynote speech delivered at the ASEAN+3 High-Level Seminar on *Fostering Bond Markets in Asia*. March, Tokyo, Japan, 30.
- Tourk, K. 2004. The political economy of east Asian economic integration. *Journal of Asian Economics* 15 (5): 843–888.
- Tran, H.Q., and J. Roldos 2003. Asian bond markets, the role of securitization and credit guarantees'. *The 2nd ASEAN Central Bank Governor's Meeting*, Bangkok, 8, 2.
- Wheatley, Jonathan. 2021. Feeble growth and debt imperil emerging economies' recovery. *Financial Times*, March 15, 2021.
- Yam, J. 1997. *Development of the debt market, keynote address at the asian debt conference*. http://www.info.gov.hk/hkma/eng/speeches/speeches/joseph/speech_140797b.htm, 14 July.

Geraldo M. Vasconcellos and Richard J. Kish

Abstract

Cross-border mergers and acquisitions have shown tremendous growth over time primarily due to a desire to circumvent tariffs and nontariff barriers arising from arms-length international trade and taxes; to obtain new options for financing; to access technology; and to distribute research and development costs over a broader base. Several factors put in place to moderate this growth include protecting key industries, limiting controlling interest levels, and restricting remittances of profits and dividends. This chapter focuses on cross-border mergers and acquisitions, and their financial and economic (both macro and micro) underpinnings, which affect their direction and magnitude. In general terms, empirical analysis supports the fact that both a host country's and the foreign country's stock and bond prices are major causal factors that influence cross-border mergers and acquisitions.

Keywords

Acquisition · Barriers · Cross-border · Diversification · International · Mergers · Multinational · Synergy · Takeovers · Tariffs · Undervaluation

One of the remarkable developments that accompanied the vigorous growth in international trade in the post-World War II era has been an unabated increase in international direct investment. This phenomenon, including its theoretical underpinnings, benefits and costs, has been the subject of voluminous research. In addition, many studies have examined the attendant questions of the host country attitudes toward international direct investment. Extant research suggests that some of the main benefits of international direct investment can be found in the avoidance

G. M. Vasconcellos (✉) · R. J. Kish
Lehigh University, Bethlehem, PA, USA
e-mail: gmv0@lehigh.edu

of tariffs and nontariff barriers to arms-length international trade, in tax incentives usually associated with efforts to attract foreign investment to a particular country or region within a country, in the ability to tap different markets for short-term and long-term capital, and in the possibility of obtaining quicker and cheaper access to superior technology, as well as the ability to spread out the output of a multinational corporation's own research and development efforts over a broader market base. On the other hand, risks and constraints affecting international direct investment include closed sectors or industries, limitations on the acquisition of a controlling interest in a foreign company, limitations on remittances of profits and dividends, limitations on cross-border mergers and acquisitions and, in some extreme cases, the possibility of expropriation (Madura et al. 1991).

The countries affiliated with the Organization for Economic Cooperation and Development (which includes all the major advanced market economies) lead this impressive growth in international direct investment. The outward direct investment flows are in general larger than the inward flows. The reason is that OECD countries invest in non-OECD countries, generally less developed ones. The inward flows of foreign direct investment (FDI) in OECD countries come almost exclusively from other OECD countries, that is to say, other major industrial countries. For example, during the 1980s, the United States was the major recipient of flows of international direct investment, followed by Europe and Canada in more modest terms. Japan was the main source of flows of international direct investment. This helps to explain why the United States gave up its position as the world's largest creditor nation to become the world's largest debtor in less than a decade. In the same period of time, Japan became one of the largest creditors. The direct investment flows, however, explain only one part of these transformations. The rest of the explanation is found in portfolio investments and their reallocations.

The acquisition of a foreign firm is one of the fastest methods of entering into a foreign market. In the late 1980s and the 1990s, this method seemed especially attractive to businesses wanting to become involved in the evolving European market. As a result, there was a surge of foreign takeovers in the European Union during this period. This demonstrated that businesses had confidence in the E.U., forming a single internal market in the long run. In fact, many of these acquisitions took place before national barriers came down. The rationale for this may be attributed in part to a growing concern that a unified Europe could translate into a more protectionist "Fortress Europe." Many foreign companies believed that the only way to participate in a unified Europe was to quickly become an insider. Acquisitions subsided after the initial surge that took place in the late 1980s, due to the creation of natural barriers to entry for outsiders. Many mergers and acquisitions were taking place within the E.U., creating larger, more efficient European businesses and effectively producing fewer opportunities for foreign companies. By the early 1990s, however, acquisitions of European firms were on the rise again due to two primary factors: (1) a need to complete the restructuring that had begun in the 1980s and that could not be done by European firms alone; and (2) regulatory changes that enabled hostile takeovers to occur more easily. But this rise in

U.S. acquisitions of E.U. companies was followed by a rise in E.U. acquisitions of U.S. firms, a cycle that seem to exist within many areas of the world economy.

International direct investment, therefore, takes place in basically two forms: *de novo* entry or mergers and acquisitions. This review focuses on cross-border mergers and acquisitions, their financial and economic underpinnings, and the factors, which affect their direction and magnitude. FDI is an integral part of the developed capital markets. The significant rise in the number of cross-border mergers and acquisitions across time warrants a better understanding of the factors affecting these activities. For example, the publicity in the 1980s surrounding foreign acquisition activity in the United States created public concern over American firms being acquired by foreign entities, leading to a significant number of studies examining the wealth effects of foreign acquisitions and capital markets factors that affect acquisition activity. Since the early 1980s, the direction of the flow of cross-border acquisitions has shifted many times. During one time period, U.S. companies were acquiring foreign firms at a higher rate than they were being acquired, but by the end of the 1990s, foreign companies reversed this direction to become the predominant acquirer again. The cycle continues to this day, although of different durations. Studies done on acquisition activity between the U.S. and Britain (Vasconcellos et al. 1990), between the U.S. and Japan (Kish and Vasconcellos 1993), between Canada and the U.S. (Vasconcellos and Kish 1996), and between the U.S. and Europe (Vasconcellos and Kish 1998) explore macroeconomic variables that contributed to this phenomenon.

43.1 Macroeconomic Factors

43.1.1 Favorable Acquisition Factors

Although there are a number of factors favoring acquisition activity, we focus on four of these factors: (1) exchange rates; (2) diversification; (3) economic conditions in the host country; and (4) technology and human resources within the acquiring firm.

43.1.1.1 Exchange Rates

One view on exchange rates revolves around the fact that while there seems to exist a relationship between exchange rates and acquisition activity, there is no evidence that a change in the exchange rate improves the position of foreign acquirers relative to their host counterparts. The argument is that when the host country's currency depreciates, the host country becomes a cheaper place for any firm to do business – foreign or domestic. Thus, the relationship between foreign acquisitions and exchange rates, contending that improved capital mobility facilitates equalized, risk-adjusted returns on international investments, is minimized. Another line of argument is that a depreciated host country's currency increases FDI in the host country's businesses. The reverse also holds true, i.e., if the host currency is strong,

there should be a pause in the foreign acquisition of host firms and an upward trend in the home country's acquisitions of foreign firms.

43.1.1.2 Diversification

Given a firm's preferred risk-return position, international diversification by way of acquisition improves the risk-return tradeoff. This reasoning is based on the assumption that the covariance of returns across economies, even within the same industries, is likely to be smaller than within a single economy. The prospective acquiring company must first decide on its desired levels of risk and return. Only then should it attempt to identify countries, industries, and specific firms, which fall within its risk class. In addition, by acquiring an ongoing foreign concern, companies may be able to circumvent tariff and nontariff barriers (i.e., quotas, voluntary restraint agreements, etc.), which attempt to protect the domestic industries and contribute to market segmentation. This action improves the risk-return tradeoff by lowering the level of unsystematic risk.

43.1.1.3 Current Economic Conditions in the Home Country

Adverse economic conditions in the home country, such as a slump, recession, or capital constraint may cause firms to concentrate on their domestic business while temporarily delaying strategic international moves. Once the economy rebounds, cross-border acquisitions are likely to again become a means for increasing demand and levels of diversification.

43.1.1.4 Acquisition of Technological and Human Resources

There are cases where a firm falls behind in the level of technological knowledge necessary to compete efficiently in its industry. If a firm is unable or unwilling to develop the required technology through research and development, it may attempt to acquire a foreign firm, which is technologically more advanced. Such an acquisition allows a firm to gain a foothold in a foreign country's market, and it may transfer the acquired technology back home, in order to strengthen its position in the domestic market. Some of the firms engaging in cross-border acquisitions are either transnational firms, or striving to become one. Transnational firms are able to behave like a local company in foreign markets, tapping into human and technological resources, while possessing the leverage of a larger, diversified entity. Indeed, this strategy provides significant diversification and allows the company to realize competencies in many markets.

43.1.2 Unfavorable Acquisition Factors

The factors discussed thus far generally tend to encourage firms to make cross-border acquisitions. In contrast, other variables that often serve to restrain cross-border movement include unavailability of information, inefficient management, monopolistic power, and government restrictions and regulations.

43.1.2.1 Unavailability of Information

The contention is that information about a prospective target firm is crucial in the decision-making process of an acquiring firm. Timely and accurate information include: current market share figures; comparisons with the competition; current sales; cash flow forecasts; and company specific strengths and weaknesses. However, foreign firms may not disclose these or other relevant figures. Thus, if the necessary information to make an accurate analysis is not available, the prospective acquiring firm may be forced to delay or discontinue its plans, even though the foreign firm appears to be an attractive target on the surface. Otherwise, failure to come up with an accurate analysis may prove harmful, or possibly devastating, to the acquiring firm. However, information effects are not always harmful, such as when the acquirer may be able to obtain information not available to other market participants.

43.1.2.2 Inefficient Management

The inefficiency argument centers on the acquiring firm being able to replace incompetent or inefficient management within the acquired firm in order to better utilize the firm's assets. The hope is that the new management will be able to increase the efficiency of the acquired firm and generate a higher return. A drawback of this action is the cost of replacing inefficient management. The negative aspects of the inefficiencies argument apply to the resistance that may materialize from the foreign managers who are left in place after the shake-up, emerging in the form of negative attitudes directed at the "outsiders" taking over the firm.

43.1.2.3 Monopolistic Power

Synergy arguments in defense of domestic or cross-border acquisitions are based on the economies of scale supposedly derived from horizontal mergers, economies of scope associated with vertical mergers, or the gains from acquiring monopolistic power. However, if monopolistic or even oligopolistic power is attained by a firm or a group of firms (a difficult position in most developed countries due to the threat of antitrust action), then entry to the industry becomes more difficult for any competitor, domestic or foreign. In addition, a monopolist is much more likely to resist a takeover. Some of the barriers to entry that make cross-border acquisitions difficult include: R&D outlays; capital expenditures necessary to establish a plant; and product differentiation, sometimes tied to large advertising expenditures.

43.1.2.4 Government Restrictions and Regulations

Most governments have some form of takeover regulations in place. In many instances, government approval is mandatory before acquisition by a foreign business can occur. In addition, government restrictions may exist on capital repatriations, dividend payouts, intra-company interest payments, and other remittances. Although these restrictions seem to be more prevalent in less developed countries, even in the developed markets, regulatory actions have been used to discourage acquisition activity. For example, the Williams's Amendment within the U.S. market increased the difficulty and costs of completing tender offers. The Tax

Reform Act of 1986 has also been cited as a factor in the increased acquisition transactions between U.S. sellers and foreign buyers. However, foreign buyers from countries with tax treaties with the host country are not subject to home taxes in repatriated earnings and, therefore, should be on equal footing with their host counterparts. Research in this area shows that most of the tax effects are industry-specific.

43.2 Microeconomic Factors

New relationships between the economic agents of different countries have come into existence with the ever-increasing globalization of markets. For example, the volume of cross-border mergers and acquisitions (M&A's) involving U.S. companies has increased in both the number of transactions and the dollar value for both net bidders and net targets. The exact motivations for cross-border M&A activity are many, including macroeconomic factors, firm-specific financial characteristics, corporate strategic moves, political motives, the possibility of a *good buy*, and/or the synergistic potential from the merged firms.

International merger and acquisition waves capture the attention of not only the business press but also of academia and policymakers. The effects of this mergermania are felt by many (i.e., managers, stockholders, intermediaries, and consumers), and the dollar amounts are considerably high. To gain a better understanding of the characteristics of firms involved in the international market for corporate control, we now focus our attention on the firm-specific financial variables of both foreign companies and the host country's companies and the role that these variables have on the probability of the acquisition.

The composition of cross-border merger and acquisitions has changed over time. Contrary to the pattern in the 1970s, we have seen an increase in the relative proportion of U.S. targets and foreign acquirers in the 1980s and 1990s, with a slowdown in the first decade of the twentieth century. Among the most important factors in the past have attracted foreign firms to the U.S. market for corporate control are: (a) growth potential and accessibility to the U.S. market; (b) availability of high technology and highly skilled labor force; (c) relative easy access to financial markets; (d) under-valuation of some companies' stock; (e) relatively limited government intervention, and (f) currency fluctuations.

43.2.1 Undervaluation

The growing web of interdependencies in the global economy has developed new relationships between economic agents of different countries. Some existing international mergers and acquisitions research focuses primarily on wealth transfers. For instance, Doukas and Travlos (1988), besides offering an excellent review of this literature, contrasts the returns to shareholders from U.S. and non-U.S.-based firms expanding into foreign markets. Conn and Connell (1990) also include an extensive literature review of mergers and acquisitions within their empirical study of wealth

transfers between the U.S. and British firms, as they expand into each other's markets.

Undervaluation revolves from the existence of product and service market imperfections that cause frictions in the global market (such as transaction costs and costs associated with barriers to entry), contributing to favor the acquisition of a company already operating. This is because the amount paid for an existing company, as compared to the replacement cost of its assets, more than compensates for the costs that could have been incurred had the foreign firm started with brand new facilities. Thus, in order to minimize the acquisition costs, foreign firms attempt to follow the same pattern of analysis as their domestic counterparts and search for undervalued and/or mismanaged companies as targets for their acquisitions. This is the basic premise of the empirical study undertaken by Gonzalez et al. (1998a, b), among others.

From the target firm's viewpoint, undervaluation is described as the likelihood of a host country's firm becoming a target increasing when the firm is perceived as being undervalued. Assuming that the takeover decision is motivated by the same stimuli that encourage firms to grow internally, a number of research studies utilize Tobin's "*q*" ratio as a predictor of takeover targets. High abnormal returns, experienced by acquirers before the merger, are consistent with a high "*q*" ratio, signaling to the companies that it is time to expand. Nevertheless, the conclusion is that the effect of the "*q*" ratio is not always significant and that these effects vary over time and across countries.

Furthermore, under the assumption that the financial market rewards well-managed firms, it is commonly interpreted that a "*q*" greater than 1 is a proxy for good management. Conversely, a ratio less than one is viewed as evidence of poor management. Thus, well-managed bidders benefit substantially from tender offers, but more so when they take over poorly managed targets. Well-managed targets benefit less from tender offers than poorly managed targets. The total takeover gain is highest for tender offers by well-managed bidders, which acquire poorly managed targets. This target undervaluation implies that there is an inverse relationship between the probability of a host country's company being acquired and the Tobin's *q*. The empirical research provides support for this view.

From the bidding firm's viewpoint, undervaluation is shown as the likelihood of a foreign firm bidding for a host country's company increasing when the firm is perceived as being overvalued. Therefore, the relationship between the ratio of market value to replacement cost of assets of foreign firms to the likelihood of these companies acquiring a host country's companies is supported (i.e., there is a positive relationship between the likelihood of a foreign firm bidding for a host country's company and the ratio of market value to replacement cost of the foreign firm). Research results show the existence of direct relation between the possibility of a foreign firm bidding for a host country's firm and the Tobin's *q* of the overseas firm.

In sum, this research empirically validates undervaluation as a predictor of M&A activity within the international setting. The results support the existence of an inverse relationship between the probability of a host country's firm becoming a

target of a foreign company and the Tobin's q ratio (i.e., undervalued host country's companies are more likely to be targets of foreign companies). This is consistent with the domestic market for corporate control.

If we relate these findings to Lang et al.'s (1989) conclusions from the domestic marketplace, then we observe positive abnormal returns for foreign companies upon the announcement of the foreign firms taking over a poorly managed host country's firms. A firm's overvaluation is proxied by a Tobin's q greater than 1. Lang et al. (1989) found positive abnormal returns when a firm with a Tobin's q greater than 1 (well-managed firm) acquired an undervalued company. Furthermore, foreign acquirers and host country's targets typically belong to the same industrial sectors. This can be interpreted as foreign companies reducing acquisition costs by acquiring undervalued firms as foreign firms trying to use their business knowhow to enhance the efficiency of the host country's targets.

Management inefficiency implies that the more inefficient is a firm's management, the greater is the probability of the firm becoming a target. Examples of variables used (in addition to the Tobin's q) to gauge management efficiency are the return on equity and sales growth. When the management is inefficient, both variables tend to show a negative relationship with the probability of an acquisition. Management inefficiency complements undervaluation reasoning. This interpretation is based on the premise that management fails to use the resources of the company up to their full potential. Thus, management inefficiency implies the existence of an inverse relationship between the ratio return on equity and growth and also the probability of the host country's company becoming a target of foreign firm. Therefore, the low return on equity and growth are manifestations of low quality management and are supported in the literature, implying that the probability that a host country's company will be taken over by a foreign firm is higher in case of greater inefficiency of the management of the domestic company.

43.2.2 Synergy Hypothesis

Much of the finance and accounting literature analyzing merger and acquisition activity is focused on the existence of synergy as a source of takeover gains within the domestic marketplace. Examples of the synergy identified that can transcend international borders include economies of scale, improved production techniques, increased market share, and more profitable use of existing assets. This suggests the existence of a direct relationship between the perceived degree of *ex-ante* synergy and the number of host country's firms acquired by foreign companies. The possibility of obtaining economies of scale, improving production techniques, increasing market share, and otherwise squeezing more profits out of existing assets are major assumptions made by the proponents of the effects from synergy. Before a merger, firms are assumed to be operating at levels of asset utilization that fall short of achieving their true potentials. Thus, the management of the bidding company could

improve the performance of both the target and the expanded firm, whether on the domestic or international level.

In an extensive literature review of the sources of gains in mergers and acquisitions, Jensen and Ruback (1983) document support for the gains to the target firms' shareholders. The basic assumption within their review is that shareholders play a passive role in any takeover activity, relying on the existence of good management who, through sound investment decisions, will be able to maximize the shareholders' wealth. The consensus within their review of studies shows that the stock price of the target firm goes up at the time surrounding the announcement date. Moreover, the majority of the empirical studies of the takeover gains rely upon event study methodology to conclude that synergy is one of the main motives behind merger and acquisition activity. But event studies are primarily a measure of the reaction of a particular economic variable (e.g., stock prices) to the event of interest (e.g., the merger or acquisition announcement) measured *ex-post*. In addition, this methodology often impairs the distinction among alternative sources of gains. In other words, this methodology is not able to identify which components of the present value of net cash flows have changed.

The fact that these studies look at the efficiency gains from mergers and acquisitions (i.e., via synergy) *ex-post* might be impairing their ability to disentangle the true gains from synergy from the existence of market imperfections. Another limitation is that the event study methodology fails to account for the long-term effects of the takeover. Therefore, it is very difficult to distinguish the real sources of gains. An alternative *ex-ante* methodology is that synergy in mergers is measured by adding the acquisition premium to the difference between replacement costs and market value of the target firm (i.e., Tobin's q).

Relying on the relationship between the merger premium and the extent that replacement costs exceeds market value, a proxy successfully used the finance literature tests the effect of synergy. To measure the existence of *ex-ante* synergy, literature relies on the relationship between market value and replacement cost of the target assets. This difference is then related to the premium paid in the takeover transaction. Thus, synergy shows up as a direct relationship between the perceived degree of *ex-ante* synergy and the number of host country's targets of foreign acquisitions. Assuming that the market for corporate control is competitive, a change in value of the firm is equal to the difference between the replacement cost of the assets of the target firm and the market value of those assets, plus the premium paid in the acquisition or merger.

Although synergy is a factor in many but not all merger activities, it is only one of the many hypotheses used to explain all merger activity. Other related merger hypotheses include management inefficiency, goodwill, and barriers to entry. For example, foreign companies often acquire a host country's companies to get around market frictions that might increase the cost of doing business in the host country. Empirical studies document the relationship between merger and acquisition activity and the presence of frictions in the market as proxied by the existence of goodwill and barriers to entry in a particular industry. Typically, the degree of goodwill and barriers to entry show a direct relationship with the probability of acquisition. One

factor used to proxy goodwill is advertising expenses. The documented trend is that the higher the proportion of advertising expenses to net sales, the larger the number of customers that have some knowledge about the product or service of the firm. Alternatively, the proportion of research and development expenses to net sales is used as a proxy for barriers to entry. The higher the proportion of research and development expenses to net sales in a particular industry, the more difficult it is to enter in the industry. Thus, there is a direct relationship between the ratio of research and development expenses to net sales of a host country's firms and the probability of these firms being a target of an overseas company. But the results are not conclusive as to what is the impact of barriers to entry and goodwill in the probability of a host country's company becoming a target. Thus, it appears that the foreign firms acquired undervalued host country's companies based on what these overseas acquirers think they can put in play to improve the operations of the host firms and not necessarily on what the host companies offer to these foreign companies in terms of reducing barriers to entry or the existence of an already established customer base.

43.2.3 Maximizing the Value of the Firm

Under the assumption that the goal of corporate managers is the maximization of shareholders' wealth, the process of cross-border mergers and acquisitions flows from the neoclassical theoretical framework of maximization of the value of the firm. If the acquisition of a host country's company is a project with a net present value larger than zero, then there is an increase in the shareholders' wealth of the acquiring company. For instance, the empirical analysis by Vasconcellos et al. (1990), using a capital budget framework, measures the feasibility of a proposed foreign acquisition. Although their research was carried out on the influence of financial variables (used in the capital budgeting process) on the difference between American acquisitions of British firms and British acquisitions of American firms, some of these findings can be generalized for all cross-border M&A activity. For example, the exchange rate has a significant positive impact on the acquisition differential. In other words, foreign firms may acquire a host country's firms because of the relatively lower foreign currency value of the host country's currency (the host country's currency was "cheap"). The Kish and Vasconcellos (1993) study of cross-border acquisitions between the United States and Japan conclude that the stock prices and the costs of debt financing are the major contemporaneous causal factors; whereas exchange rates only had significance as a predictor of trends in acquisitions. Thus, generalities to fit all situations do not appear to exist. Most of the companies involved in cross-border M&As establish a sort of acquisition screening. This screening process involves country-specific and firm-specific screening variables (i.e., per capita GDP, market share of the target, etc.). The general conclusion is that the internationalization of the firm is a value-enhancing phenomenon.

The net present value (NPV) analysis assumes that the managers of the foreign firms bidding for the host country's companies decide to make the acquisition only when the decision has a positive impact on the shareholders' wealth of the foreign

company. The net present value criteria assume a positive relationship between the factors affecting the NPV criterion and the likelihood of a foreign firm acquiring a host country's company. Another frequently argued view is that a relatively large and stable ("mature") host country's companies are more likely to go overseas than the average host's firms. Thus, the mature firm argument states that a host country's bidders in the cross-border merger and acquisition market are more likely to be mature firms.

Cross border M&A research start from the assumption that in the international market for corporate control, firms decide about an acquisition project using essentially the same decision-making framework that the firms would use for internal projects. Research supports the net present value approach and the assumption that the management of the foreign firm will undertake projects that have a positive impact on the wealth of its shareholders. The empirical research shows the existence of a positive relationship between the factors affecting the NPV criterion and the likelihood of a foreign firm acquiring a host country's firms.

Foreign firms also seem to be more likely to acquire a host country's companies with high debt capacity. A substantial debt capacity can be utilized to reduce the cost of the acquisition through debt financing at relatively low cost, whereas a high debt to equity ratio could increase the cost of new debt financing. Foreign firms are more likely to acquire a host country's companies with relatively high liquidity, as evidenced by the importance of the current ratio in the literature. In addition, host country's companies with relatively low price of stock to earnings are more probable to be acquired, serving as evidence that the managers of foreign companies acquiring host country's companies make their merger and/or acquisition decisions pursuing the maximization of the foreign companies shareholders' wealth.

In addition to examining the financial characteristics of the host country's targets from cross-border M&A, the same analysis for the host country's bidders in the global takeover market has been summarized. The reasons for a host country's FDI have been widely discussed in the literature. The motivations leading to host country's FDI include product market imperfections, institutional imperfections (i.e., differentials in tax laws), and limitations of the domestic market.

Jensen (1988) argues that firms with free cash flows will be likely bidders in the takeover market. Thus, mature firms in a host country are more likely to be bidders in the cross-border M&A market. Normally, a company follows a life cycle that is closely connected to product line development. A mature firm has a relatively stable financial profile and may face two options: to become "better" or to get "bigger" In order to become "bigger," these companies may attempt to go overseas.

The following financial variables proxy for identifying mature companies relative to the industry: net sales growth, size of total assets, price-earnings ratio, and free cash flow. There is an inverse relationship between both the growth and the price-earnings ratio of a host country's firm and the probability of this firm becoming a bidder in the global market for corporate control. Furthermore, there is direct association between both the size and free cash flow of a host country's company relative to the industry and the likelihood of this company becoming a bidder for a foreign firm. In addition, foreign firms with a "Tobin's q " greater than 1 are more

likely to acquire a host country's companies. This is consistent with Jensen's (1988) conclusions. Relatively high "Tobin's q " firms may have enough resources to invest in the acquisition of other firms. The exchange rate does not have a strong impact on the probability of acquisition of a host country's company. For example, a very strong dollar during the first half of the 1980s and a weak dollar the second half failed to impact the number of U.S. companies acquired or acquiring in that they were on average the same. There are alternative (and not mutually exclusive) explanations for the difference on the importance attributed to the exchange rate. First, most of the studies found in the literature examining the difference between the number of host country's acquisitions of foreign companies versus the number of foreign acquisitions of host country's firms report inconsistent results. Second, the exchange rate could affect the timing of the acquisition but not the acquisition decision itself. The other possibility is that there are different time periods being studied. Also found was that the foreign firms have a relatively high return on equity when compared to the industry average. Since return on equity is used as a proxy for management efficiency, the conclusion is that foreign companies with above average efficiency in their countries have a higher likelihood of acquiring a host country's firms.

The combined results on "Tobin's q " for the host country's targets and foreign bidders mirror the domestic case of mergers and acquisitions. That is, high "Tobin's q " foreign bidders had positive abnormal returns when they acquired targets with "Tobin's q " < 1 . Research supports the share-holders' wealth maximization theory as applied to the investment decision of whether or not to acquire a host country's companies. Finally, the host country's companies going overseas are "mature" companies with large amounts of assets, considerable free cash flows, and low growth. The fact that the host country's companies acquiring foreign companies have an average low price to earnings ratio may be interpreted as a move of the management of these host country's companies to attempt to maximize its share-holders' wealth by signaling to the market that the increase in globalization of the company's operations is a risk reduction event due to diversification.

43.3 An Analytical View of Cross-Border Mergers and Acquisitions

The feasibility of a foreign acquisition can be evaluated first like any other project, with specific attention to peculiar characteristics. Capital budgeting analysis can be applied to determine whether the NPV of the acquisition is positive. Consider the following capital budgeting framework, as applied to a foreign acquisition:

$$NPV_{FA} = -I_{FA} + \sum_{t=1}^n \frac{CF_{FA,t}}{(1 + k_{FA})^t} + \frac{SV_{FA,n}}{(1 + k_{FA})^n} \quad (1)$$

Where NPV_{FA} is the net present value of a foreign acquisition; I_{FA} the initial outlay of a foreign acquisition; k_{FA} the required return on the foreign acquisition; CF_{FA} the

cash flows to the acquirer; SV the salvage value to the acquirer; t the = time period and n the number of periods in which the project is expected to exist.

As with any project, the variables above should incorporate any tax implications so that the net present value reflects after-tax cash flows. In addition, all cash flows should be measured from the acquirer's perspective and in the acquirer's home currency.

Breaking the general NPV equation into its components can identify the factors that influence a firm's attraction to a prospective foreign acquirer. The following discussion identifies the specific factors, which affect a foreign acquisition's initial outlay, periodic cash flows, and salvage value. The initial outlay (I_{FA}) can be broken down into three components, as shown below:

$$I_{FA} = E_h + D_h + D_f(ER_f) \quad (2)$$

where E_h the equity in the home currency; D_h borrowed funds in the home currency; D_f the borrowed funds in the foreign currency; and ER_f the exchange rate of foreign currency at the time the foreign funds were borrowed.

To measure the entire initial outlay in terms of the home currency, any foreign funds borrowed by the acquiring firm must be translated into the home currency. Moreover, some firms may cover the entire initial outlay from any one of the above components.

The relevant cash flows in the analysis of cross-border mergers and acquisitions are those received by the acquiring firm. These cash flows are determined by: (1) the after-tax foreign cash flows generated; (2) the percentage of those after-tax cash flows to be remitted to the acquirer; and (3) the exchange rates at the time the after-tax foreign cash flows are remitted. Then, the after-tax cash flows received by the acquiring firm can be described as:

$$CF_{FA,t} = (CF_{f,t})(1 - R_{f,t})(ER_{f,t}) \quad (3)$$

where $CF_{f,t}$ is the foreign cash flows generated during period t ; $R_{f,t}$ the proportion of cash flows retained by the (then) foreign subsidiary to support future operations; and $ER_{f,t}$ the exchange rate of the foreign currency at the time cash flows are remitted to the acquiring firm.

The salvage value from the acquirer's perspective as of time n ($SV_{FA,n}$) is determined by the anticipated foreign market value of the acquired business at time n ($MV_{f,n}$), and the prevailing exchange rate at the time of the planned sale, as described below:

$$SV_{FA,n} = (MV_{f,n})(ER_{f,n}) \quad (4)$$

Note that the foreign value may represent a liquidation value or a going concern value, whichever is likely to be higher.

Integrating the detailed expressions for the initial outlay, periodic cash flows, and salvage value, a comprehensive expression for the NPV analysis of a foreign acquisition can be written as follows:

$$\begin{aligned}
 NPV_{FA} &= -I_{FA} + \sum_{t=1}^n \frac{CF_{FA,t}}{(1+k_{FA})^t} + \frac{SV_{FA,n}}{(1+k_{FA})^n} \\
 &= -[E_h + D_h + D_f(ER_f)] \\
 &\quad + \sum_{t=1}^n \frac{[(CF_{f,t})(1-R_{f,t})(ER_{f,t})]}{(1+k_{FA})^t} \frac{[(MV_{f,n})(ER_{f,n})]}{(1+k_{FA})^n}
 \end{aligned} \tag{5}$$

When expressed as in Eq. 5, the capital budgeting approach provides a valuable framework for explaining the influence of several factors regarding the feasibility of foreign acquisitions.

In conclusion, the phenomenon of cross-border mergers and acquisitions has shown vitality in the last two decades and the trend appears set to continue in the new century. For example, the UNCTAD's *World Investment Report, 2000* reported that the overall value of the flow of cross-border mergers and acquisitions was \$151 billion in 1991 and increased to \$720 billion in 1999 (UNCTAD 2000). In addition, the annual growth rates of these flows are shown to be 26.4% for 1986–1990, 23.3% for 1991–1995, and 46.9% for 1996–1999. Moreover, the quickly evolving single European market in the late 1980s and early 1990s encouraged many non-European firms to establish a presence in Europe before the barriers to entry intensified. Consequently, by the mid-1990s U.S. FDIs in the European Union increased by approximately 200% from the early 1980s. In general terms, empirical analysis supports the fact that both a host country's and the foreign country's stock prices are a major causal factor that influence cross-border mergers and acquisitions. Bond yields are also shown to be major causal factors. This implies that bond yields may be one of the final negotiating points in the decision to consummate an acquisition. Finally, the exchange rate does not consistently acquire significance for all countries. Thus, the exchange rate can only serve as a predictor of trends in acquisitions.

References

- Conn, R.L., and F. Connell. 1990. International mergers: Returns to U.S. and British firms. *Journal of Business Finance and Accounting* 17 (5): 689–712.
- Doukas, J., and N.G. Travlos. 1988. The effect of corporate multinationalism on shareholders' wealth: Evidence from international acquisitions. *Journal of Finance* 43: 1161–1175.
- Gonzalez, P., G.M. Vasconcellos, and R.J. Kish. 1998a. Cross-border mergers and acquisitions: The undervaluation hypothesis. *The Quarterly Review of Economics and Finance* 38 (1): 25–45.
- . 1998b. Cross-border mergers and acquisitions: The synergy hypothesis. *The International Journal of Finance* 10 (4): 1297–1319.
- Jenson, M. 1988. Takeovers: Their causes and consequences. *The Journal of Economic Perspectives* 2 (1): 21–48.
- Jensen, M., and R.S. Ruback. 1983. The market for corporate control the scientific evidence. *Journal of Financial Economics* 11: 5–50.
- Kish, R.J., and G.M. Vasconcellos. 1993. An empirical analysis of factors affecting cross-border acquisitions: US-Japan. *Management International Review* 33 (3): 222–245.

- Lang, L.H., R.M. Stulz, and R.A. Walking. 1989. Managerial performance, Tobin's Q, and the gain from successful tender offers. *Journal of Financial Economics* 24: 137–154.
- Madura, J., G.M. Vasconcellos, and R.J. Kish. 1991. A valuation model for international acquisitions. *Management Decision* 29 (4): 31–38.
- United Nations Conference on Trade and Development (UNCTAD). 2000. *World Investment Report 2000: Cross-Border mergers and acquisitions and development*. Switzerland: New York and Geneva.
- Vasconcellos, G.M., and R.J. Kish. 1996. Factors affecting cross-border mergers and acquisitions: The Canada–US experience. *Global Finance Journal* 7 (2): 223–238.
- . 1998. Cross-border mergers and acquisitions: The European–US experience. *Journal of Multinational Financial Management* 8: 431–450.
- Vasconcellos, G.M., J. Madura, and R.J. Kish. 1990. An empirical investigation of factors affecting cross-border acquisitions: The United States vs. United Kingdom experience. *Global Finance Journal* 1 (3): 173–189.

Shin-Huei Wang

Abstract

Jump diffusion processes have been used in modern finance to capture discontinuous behavior in asset pricing. Various jump diffusion models are considered in this chapter. Also, the applications of jump diffusion processes on stocks, bonds, and interest rate are discussed.

Keywords

Affine jump diffusion model · ARCH/GARCH jump diffusion model · Autoregressive jump process model · Bernoulli jump process · Black–Scholes model · Conditional jump dynamics · Gauss–Hermite jump process · Jump diffusion process · Jump diffusion with conditional heteroskedasticity · Mixed-jump process

44.1 Introduction

In contrast to basic insights into continuous-time asset-pricing models that have been driven by stochastic diffusion processes with continuous sample paths, jump diffusion processes have been used in finance to capture discontinuous behavior in asset pricing. As described in Merton (1976), the validity of Black–Scholes formula depends on whether the stock price dynamics can be described by a continuous-time diffusion process whose sample path is continuous with probability 1. Thus, if the stock price dynamics cannot be represented by stochastic process with a continuous sample path, the Black–Scholes solution is not valid. In other words, as the price processes feature big

S.-H. Wang (✉)

University of Southern California, Los Angeles, CA, USA

The Center for Operations Research and Econometrics (CORE), Louvain-la-Neuve, Belgium

e-mail: wang@core.ucl.ac.be

jumps, i.e., not continuous, continuous-time models cannot explain why the jumps occur, and hence not adequate. In addition, Ahn and Thompson (1986) also examined the effect of regulatory risks on the valuation of public utilities and found that those “jump risks” were priced even though they were uncorrelated with market factors. It shows that jump risks cannot be ignored in the pricing of assets. Thus, a “jump” stochastic process defined in continuous time, and also called as “jump diffusion model” was rapidly developed.

The jump diffusion process is based on Poisson process, which can be used for modeling systematic jumps caused by surprise effect. Suppose we observe a stochastic process S_t , which satisfies the following stochastic differential equation with jump:

$$dS_t = a_t dt + \sigma_t dW_t + dJ_t, \quad t \geq 0 \quad (1)$$

where dW_t is a standard Wiener process. The term dJ_t represents possible unanticipated jumps, and which is a Poisson process. As defined in Gouriéroux and Jasiak (2001), a jump process $(J_t, t \in R^+)$ is an increasing process such that

- (i) $J_0 = 0$
- (ii) $P[J_{t+dt} - J_t = 1 | J_t] = \lambda_t dt + o(dt),$
- (iii) $P[J_{t+dt} - J_t = 0 | J_t] = 1 - \lambda_t dt + o(dt),$

where $o(dt)$ tends to 0 when t tends to 0, and λ_t , called the intensity, is a function of the information available at time t . Furthermore, since the term dJ_t is part of the unpredictable innovation terms we make $E[\Delta J_t] = 0$, which has zero mean during a finite interval h . Besides, as any predictable part of the jumps may be included in the drift component α_t , jump times $\tau_{ij} = 1, 2, \dots$ vary by some discrete and random amount. Without loss of generality, we assume that there are k possible types of jumps, with size a_i , $i = 1, 2, L$, and the jumps occur at rate λ_t that may depend on the latest observed S_t . As soon as a jump occurs, the jump type is selected randomly and independently. The probability of a jump of size a_i , occurring is given by p_i . Particularly, for the case of the standard Poisson process, all jumps have size 1. In short, the path of a jump process is an increasing stepwise function with jumps equal to 1 at random rate $D_1, D_2, \dots, D_b, \dots$

Related research on the earlier development of a basic Poisson jump model in finance was by Press (1967). His model can be motivated as the aggregation of a number of price changes within a fixed-time interval. In his paper, the Poisson distribution governs the number of events that result in price movement, and the average number of events in a time interval is called intensity. In addition, he assumes that all volatility dynamics is the result of discrete jumps in stock returns and the size of a jump is stochastic and normally distributed. Consequently, some empirical

applications found that a normal Poisson jump model provides a good statistical characterization of daily exchange rate and stock returns. For instance, using Standard and Poor's 500 futures options and assuming an underlying jump diffusion, Bates (1991) found systematic behavior in expected jumps before the 1987 stock market crash. In practice, by observing different paths of asset prices with respect to different assets, distinct jump diffusion models were introduced into literature by many researchers. Therefore, in this chapter, we will survey various jump diffusion models in current literature as well as estimation procedures for these processes.

44.2 Mixed-Jump Processes

The total change in asset prices may be comprised of two types of changes:

Normal vibrations caused by marginal information events satisfying a local Markov property and modeled by a standard geometric Brownian motion with a constant variance per unit time. It has a continuous sample path.

Abnormal vibrations caused by information shocks satisfying an antipathetical jump process defined in continuous time, and modeled by a jump process, reflecting the nonmarginal impact of the information.

Thus, there have been a variety of studies that explain too many outliers for a simple, constant-variance log-normal distribution of stock price series. Among them, Merton (1976) and Tucker and Pond (1988) provide a more thorough discussion of mixed-jump processes. Mixed-jump processes are formed by combining a continuous diffusion process and a discrete-jump process and may capture local and nonlocal asset price dynamics.

Merton (1976) pioneered the use of jump processes in continuous-time finance. He derived an option pricing formula as the underlying stock returns are generated by a mixture of both continuous and the jump processes. He posited stock returns as

$$\frac{dS}{S} = (\alpha - \lambda k)dt + \sigma dZ + dq \quad (2)$$

where S is the stock price, α the instantaneous expected return on the stock, σ^2 the instantaneous variance of the stock return conditional on no arrivals of "abnormal" information, dZ the standardized Wiener process, q the Poisson process assumed independent of dZ , λ the intensity of the Poisson process, $k = \varepsilon(Y - 1)$, where $\tilde{Y} - 1$ is the random variable percentage change in stock price if the Poisson event occurs; ε is the expectation operator over the random variable Y . Actually, Eq. 1 can be rewritten as

$$\begin{aligned} \frac{dS}{S} &= (\alpha - \lambda k)dt + \sigma dZ && \text{if the Poisson does not occur,} \\ &= (\alpha - \lambda k)dt + \sigma dZ + (Y - 1), && \text{if the Poisson occurs} \end{aligned} \quad (3)$$

Therefore, the option return dynamics can be rewritten as

$$\frac{dW}{W} = (\alpha_w - \lambda k_w)dt + \sigma_w dZ + dq_w \quad (4)$$

Most likely, α_w is the instantaneous expected return on the option, σ_w^2 is the instantaneous variance of the stock return conditional on no arrivals of “abnormal” information, q_w is an Poisson process with parameter λ assumed independent of dZ , $k_w = \varepsilon(Y_w - 1)$, $(Y_w - 1)$ is the random variable percentage change in option price if the Poisson event occurs, ε is the expectation operator over the random variable Y_w . The Poisson event for the option price occurs if and only if the Poisson event for the stock price occurs. Further, define the random variable, X_n , which has the same distribution as the product of n independently and identically distributed random variables. Each of n independently and identically distributed random variables has the identical distribution as the random variable Y described in Eq. 1. As a consequence, by the original Black–Scholes option pricing formula for the no-jump case, $W(S, \tau; E, r, \sigma^2)$, we can get the option price with jump component

$$F(s, \tau) = \sum_{n=1}^{\infty} \frac{e^{-\lambda\tau(\lambda\tau)^n}}{n!} [\varepsilon_n \{W(SX_n e^{-\lambda k\tau, \tau}; E, \sigma^2, r)\}] \quad (5)$$

Generally speaking, W satisfies the boundary conditions of partial differential equation (see Oksendal 1998), and can be rewritten as a twice continuously differentiable function of the stock price and time, $W(t) = F(S, t)$. Nevertheless, Eq. 5 still not only holds most of the attractive features of the original Black–Scholes formula such as being regardless of the investor preferences or knowledge of the expected return on the underlying stock, but also satisfies the Sharpe–Lintner Capital Asset Pricing model as long as the jump component of a security’s return is uncorrelated with the market. In other words, the mixed-jump model of Merton uses the CAPM to value options written on securities involving jump processes.

Also, Tucker and Pond (1988) empirically investigated four candidate processes (the scaled- t distribution, the general stable distribution, compound normal distribution, and the mixed-jump model) for characterizing daily exchange rate changes for six major trading currencies from the period 1980 to 1984. They found that the mixed-jump model exhibited the best distributional fit for all six currencies tested. Akgiray and Booth (1988) also found that the mixed-diffusion jump process was superior to the stable laws or mixture of normals as a model of exchange rate changes for the British pound, French franc, and the West German mark relative to the U.S. dollar. Thus, both theoretical and empirical studies of exchange rate theories under uncertainty should explicitly allow for the presence of discontinuities in exchange rate processes. In addition, the assumption of pure diffusion processes for exchange rates could lead to misleading inferences due to its crude approximation.

44.3 Bernoulli Jump Process

In the implementation of empirical works, Ball and Torous (1983) provide statistical evidence with the existence of log-normally distributed jumps in a majority of the daily returns of a sample of NYSE-listed common stocks. The expression of their Poisson jump diffusion model is as Eq. 1, and jump size Y has posited distribution, $\ln Y \sim N(\mu, \delta^2)$.

Ball and Torous (1983) introduced the Bernoulli jump process as an appropriate model for stock price jumps. Denote X_i as the number of events that occur in subinterval i and independent distributed random variables. By stationary independent increment assumption,

$$N = \sum_{i=1}^n X_i$$

where N is the number of events that occur in a time interval of length t . Besides, define $h = t/n$ for any arbitrary integer n and divide $(0, t)$ into n equal subintervals each of length h . Thus, X_i satisfies

$$\begin{aligned}\Pr[X_i = 0] &= 1 - \lambda h + O(h) \\ \Pr[X_i = 1] &= \lambda h + O(h) \text{ for } i = 1, 2, \dots, n \\ \Pr[X_i > 1] &= O(h)\end{aligned}$$

For large n , X_i has approximately the Bernoulli distribution with parameter $\lambda h = \lambda t/n$. As a result, N has the binomial distribution, approximately, i.e.,

$$\Pr[N = k] \cong \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \quad k = 0, 1, 2, \dots, n.$$

Now, assume that t is very small, they can approximate N by the Bernoulli variate X defined by

$$\begin{aligned}\Pr[X = 0] &= 1 - \lambda t, \\ \Pr[X = 1] &= \lambda t.\end{aligned}$$

The advantage of the Bernoulli $\neq$ jump process is that more satisfactory empirical analyses are available. The maximum likelihood estimation can be practically implemented and the unbiased, consistent, and efficient estimators that attain the Cramer–Rao lower bound for the corresponding parameters. Moreover, the statistically most powerful test of the null hypothesis $\lambda = 0$ can be implemented. Obviously, a Bernoulli jump process models information arrivals and stock price jumps. This shows that the presence of a jump component in common stock returns can be

possessed well. As a consequence, Vlaar and Palm (1993) combined the GARCH (1,1) and Bernoulli jump distribution to account for skewness and leptokurtosis for weekly rates of the European Monetary System (EMS). Das (2002) considered the concept of Bernoulli approximation to test the impact of Federal Reserve actions by Federal Funds' rate as well (see Sects. 43.9.2 and 43.5, respectively).

44.4 Gauss–Hermite Jump Process

To ensure the efficiency properties in valuing compound option, Omberg (1988) derived a family of jump models by employing Gauss–Hermite quadrature.

Note that $t = 0$ and $t = T$ are the current time and expiration date of the option, respectively, and $\Delta t = T/N$. Consider a compound option that can only be exercised at the N interval boundaries $t_k = T - k\Delta t$, $k = 0, \dots, N$. Let $C_k(S)$ be the value of the compound option at time t_k , the current value of the compound option is then $C_N(S)$; the value of an actual contingent claim with optimal exercise possible at any time is $\lim N \rightarrow C_N(S)$. The compound option can be recursively valued by

$$C_{k+1}(S) = \max \{EV_{k+1}, e^{-r} \Delta t E[C_k(S_k; S)]\},$$

where EV_{k+1} is the immediate exercise value at time t_{k+1} . Since $S(t)$ is an unrestricted log-normal diffusion process from t_k to t_{k+1} ,

$$E[C_k] = \int_{-\infty}^{\infty} \phi(z) C_k \left(S e^{\mu' \Delta t + z \sigma \sqrt{\Delta t}} \right) dz, \quad (6)$$

where z is an independent sample from a normal distribution with mean zero and variance one, $\phi(z)$ is its density function, and $\mu' = r - \sigma^2/2$ for a risk-neutral valuation. A jump process approximation to the above with n jumps takes the form

$$E[C_k] \cong \sum_{j=1}^n p_j C_k(S e^{\mu' j}), p_j \leq 0 \text{ for } j = 1, \dots, n,$$

$$\sum_{j=1}^n p_j = 1, u_j = \mu' \Delta t + z_j \sigma \sqrt{\Delta t}$$

So, Omberg (1988) considers to use Gaussian integration to approximate an integration of the form as in Eq. 6. For example, for the integral,

$$I = \int_a^b w(x) f(x) dx,$$

we can approximate this equation by a weighted average of the function $f(x)$ at n points $\{x_1, \dots, x_n\}$. Let $\{w_i\}$ and $\{x_j\}$ are selected to maximize the degree of precision m^* , which is a integration rule, i.e., if the integration error is zero for all

polynomials $f(x)$ of order m^* or less. $\{P_j(x)\}$ is the set of polynomials with respect to the weighting function $w(x)$,

$$\begin{aligned}\int_a^b w(x)P_i(x)P_j(x)dx &= 0, \text{ for } i \neq j, \\ \int_a^b w(x)P_j^2(x)dx &= \gamma_j \neq 0, \text{ for } i = j,\end{aligned}$$

Thus, the optimal evaluation points $\{X_j\}$ are the n zeros of $P_n(x)$ and the corresponding weights $\{w_j\}$ are

$$w_j = \frac{-(a_{n+1,n+1}/a_{n,n})\gamma_n}{P'_n(x_j)P_{n+1}(x_j)} > 0.$$

The degree of precision is $m^* = 2n - 1$. If the weighting function $w(x)$ is symmetric with regard to the midpoint of the interval $[a, b]$, then $\{x_j\}$ and $\{w_j\}$ are the Gaussian evaluation points $\{x_j\}$ and weights $\{w_j\}$, respectively. Particularly, the above procedure is called Gauss–Hermite quadrature to approximate the integration problem. What is shown in Omberg (1988) is the application of Gauss–Hermite quadrature to the valuation of a compound option, which is a natural way to generate jump processes of any order n that are efficient in option valuation. Thus, the Gauss–Hermite jump process arises as an efficient solution to the problem of replicating a contingent claim over a finite period of time with a portfolio of assets. With this result, he suggested the extension of these methods to option valuation problems with multiple state variables, such as the valuation of bond options in which the state variables are taken to be interest rates at various terms.

44.5 Jumps in Interest Rates

Cox et al. (1985a) proposed an influential paper that derived a general equilibrium asset pricing model under the assumption of diffusion processes, and analyzed the term structure of interest rate by it. Ahn and Thompson (1988) applied Cox, Ingersoll, and Ross's methodology to their model, which is driven by jump diffusion processes, and investigated the effect of jump components of the underlying processes on the term structure of interest rates. They differ from the model of Cox et al. (1985b) when they consider the state variables as jump diffusion processes. Therefore, they suggested that jump risks may have important implications for interest rate, and cannot be ignored for the pricing of assets. In other words, they found that Merton's (1973) multi-beta CAPM does not hold in general due to the existence of jump component of the underlying processes on the term structure of interest rate. Also, Breeden's single consumption beta does not hold, because the discontinuous movements of the investment opportunities cannot be fully captured by a single consumption beta. Moreover, in contrast with the work of Cox et al. (1981) providing that the traditional expectations theory is not consistent with the equilibrium

models, they found that traditional expectations theory is not consistent with the equilibrium models as the term structure of interest rate is under the jump diffusion process, since the term premium is affected by the jump risk premiums. Das (2002) tested the impact of Federal Reserve actions by examining the role of jump-enhanced stochastic processes in modeling the Federal Funds rate. This research illustrated that compared to the stochastic processes of equities and foreign exchange rates, the analytics for interest rates are more complicated. One source of analytical complexity considered in modeling interest rates with jumps is mean reversion. Allowing for mean reversion included in jump diffusion processes, the process for interest rates employed in that paper is as follows

$$dr = k(\theta - r)dt + vdz + Jd\pi(h), \quad (7)$$

which shows interest rate has mean-reversing drift and two random terms, a pure diffusion process and a Poisson process with a random jump J . In addition, the variance of the diffusion is v^2 , and a Poisson process π represents the arrival of jumps with arrival frequency parameter h , which is defined as the number of jumps per year. Moreover, denote J as jump size, which can be a constant or with a probability distribution. The diffusion and Poisson processes are independent of each other as well as independent of J .

The estimation method used here is the Bernoulli approximation proposed in Ball and Torous (1983). Assuming that there exists no jump or only one jump in each time interval, approximate the likelihood function for the Poisson–Gauss model using a Bernoulli mixture of the normal distributions governing the diffusion and jump shocks.

In discrete time, Eq. 7 can be expressed as follows:

$$\Delta r = k(\theta - r)\Delta t + v\Delta z + J(\mu, \gamma^2)\Delta\pi(h),$$

where v^2 is the annualized variance of the Gaussian shock, and Δz is a standard normal shock term. $J(\mu, \gamma^2)$ is the jump shock with normal distribution. $\Delta\pi(q)$ is the discrete-time Poisson increment, approximated by a Bernoulli distribution with parameter $q = h\Delta t + O(\Delta t)$, allowing the jump intensity q to depend on various state variables conditionally. The transition probabilities for interest rates following a Poisson–Gaussian process are written as (for $s > t$):

$$\begin{aligned} f[r(s)r(t)] = & q \exp\left(\frac{-(r(s) - r(t) - k(\theta - r(t))\Delta t - \mu)^2}{2(v_t^2\Delta t + \gamma^2)}\right) \frac{1}{\sqrt{(2\pi(v_t^2\Delta t + \gamma^2))}} \\ & + (1 - q) \exp\left(\frac{-(r(s) - r(t) - k(\theta - r(t))\Delta t)^2}{2v_t^2\Delta t}\right) \frac{1}{\sqrt{(\pi v_t^2\Delta t)}} \end{aligned}$$

where $q = h\Delta t + O(\Delta t)$. This is an approximation for the true Poisson–Gaussian density with a mixture of normal distributions. As in Ball and Torous (1983, 1985), by maximum-likelihood estimation, which maximizes the following function L ,

$$L = \prod_{t=1}^T f[r(t + \Delta t)|r(t)],$$

we can obtain estimates that are consistent, unbiased, and efficient and attain the Cramer-Rao lower bound. Thus, they obtain the evidence that jumps are an essential component of interest rate models. Especially, the addition of a jump process diminishes the extent of nonlinearity although some research finds that the drift term in the stochastic process for interest rates appears to be nonlinear.

Johannes (2004) suggested the estimated infinitesimal conditional moments to examine the statistical and economic role of jumps in continuous-time interest rate models. Based on Johannes's approach, Bandi and Nguyen (2003) provided a general asymptotic theory for the full function estimates of the infinitesimal moments of continuous-time models with discontinuous sample paths of the jump diffusion type. Their framework justifies consistent nonparametric extraction of the parameters and functions that drive the dynamic evolution of the process of interest. (i.e., the potentially nonaffine and level dependent intensity of the jump arrival being an example). Particularly, Singleton (2001) provided characteristic function approaches to deal with the Affine jump diffusion models of interest rate. In the next section, we will introduce affine jump diffusion model.

44.6 Affine Jump Diffusion Model

For development in dynamic asset pricing models, a particular assumption is that the state vector X follows an affine jump diffusion (AJD). An affine jump model is a jump diffusion process. In general, as defined in Duffie and Kan (1996), we suppose the diffusion for a Markov process X is 'affine' if

$$\begin{aligned}\mu(y) &= \theta + \kappa y \\ \sigma(y)\sigma(y)' &= h + \sum_{j=1}^N y_j H^{(j)},\end{aligned}$$

where $\mu : D \rightarrow R^n$ and $\sigma : D \rightarrow R^n \times n$, θ is $N \times 1$, κ is $N \times N$, h and $H^{(j)}$ are all $N \times N$ and symmetric. The X 's may represent observed asset returns or prices or unobserved state variables in a dynamic pricing model, such as affine term structure models. Thus, extending the concept of 'affine' to the case of affine jump diffusions, we can note that the properties for affine jump diffusions are that the drift vector, 'instantaneous' covariance matrix, and jump intensities all have affine dependence on the state vector. Vasicek (1977) and Cox et al. (1985b) proposed the Gaussian and square root diffusion models which are among the AJD models in term structure literature. Suppose that X is a Markov process in some state space $D \subset R^n$, the affine jump diffusion is

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dZ_t,$$

where W is an standard Brownian motion in R^n , $\mu : D \rightarrow R^n$, $\sigma : D \rightarrow R^{n \times n}$, and Z is a pure jump process whose jumps have a fixed probability distribution ν on and arriving intensity $\{\lambda(X_t) : t \geq 0\}$, for some $\lambda : D \rightarrow [0, \infty)$.

Furthermore, in Duffie et al. (2000), they suppose that X is Markov process whose transition semi-group has an infinitesimal generator of levy type defined at a bounded C^2 function $f : D \rightarrow R$ with bounded first and second derives by

$$\mathcal{G}f(x) = f_x(x)\mu(x) + \frac{1}{2}tr\left[f_{xx}(x)\sigma(x)\sigma(x)^T\right] + \lambda(x).$$

It means that conditional on the path of X , the jump times of Z are the jump times of a Poisson process with time varying intensity $\{\lambda(X_s) : 0 \leq s \leq t\}$, and that the size of the jump of Z at a jump time T is independent of $\{X_s : 0 \leq s \leq T\}$, and has the probability distribution ν . Consequently, they provide an analytical treatment of a class of transforms, including Laplace and Fourier transformations in the setting of affine jump diffusion state process.

The first step to their method is to show that the Fourier transform of X_t and of certain related random variables are known in closed form. Next, by inverting this transform, they show how the distribution of X_t and the prices of options can be recovered. Then, they fix an affine discount rate function $R : D \rightarrow R$. Depending on coefficients (K, H, L, ρ) , the affine dependence of μ , $\sigma\sigma^T$, λ , R are determined, as shown in p. 1350 of Duffie et al. (2000). Moreover, for $c \in C^n$, the set of n -tuples of complex numbers, let $\theta(c) = \int R^n \exp(c \cdot z) d\nu(z)$. Thus, the “jump transform” θ determines the jump size distribution. In other words, the “coefficients” (K, H, L, θ) of X completely determine its distribution. Their method suggests a real advantage of choosing a jump distribution ν with an explicitly known or easily computed jump transform θ . They also applied their transform analysis to the pricing of options (see Duffie et al. 2000). Furthermore, Singleton (2001) developed several estimation strategies for affine asset pricing models based on the known functional form of the conditional characteristic function (CCF) of discretely sampled observations from an affine jump diffusion model, such as LML-CCF (Limited-information estimation), ML-CCF (Maximum likelihood estimation), and GMM-CCF estimation, etc. As shown in his paper, a method of moments estimator based on the CCF is shown to approximate the efficiency of maximum likelihood for affine diffusion models.

44.7 Geometric Jump Diffusion Model

Using Geometric Jump Diffusion with the instantaneous conditional variance, V_t , following a mean reverting square root process, Bates (1996) showed that the exchange rate, $S(\$/\text{deutschemark (DM)})$ followed it:

$$\begin{aligned}
dS/S &= (\mu - \lambda \bar{k})dt + \sqrt{V}dZ + k dq \\
dV &= (\alpha - \beta V)dt + \sigma_v \sqrt{V}dZ_v \\
\text{Cov}(dZ, dZ_v) &= \rho dt \\
\Pr(dq = 1) &= \lambda dt \\
\ln(1 + k) &\sim N\left(\ln(1 + \bar{k}) - \frac{1}{2}\delta^2, \delta^2\right),
\end{aligned}$$

where μ is the instantaneous expected rate of appreciation of the foreign currency, λ is the numbers of jumps in a year, k is the random percentage jump conditional on a jump occurring, and q is a Poisson counter with intensity λ .

The main idea of this model illustrated that skewed distribution can arise by considering nonzero average jumps. Similarly, it also discusses that excess kurtosis can arise from a substantial jump component. In addition, this geometric jump diffusion model can see a direct relationship between the magnitude of conditional skewness and excess kurtosis and the length of the holding period as well.

44.8 Autoregressive Jump Process Model

A theory of the distribution of stock returns was derived by Bachelier (1900) and expanded using the idea of Brownian motion by Osborne (1959). However, the empirical works generally concluded that the B-O model fits observed returns rather poorly. For example, a casual examination of transactions data shows that assumption of a constant interval between transactions is not strictly valid. On the other hand, transactions for a given stock occur at random times throughout a day which gives nonuniform time intervals. Also, the notion of independence between transaction returns is suspect. Niederhoffer and Osborne (1966) showed that the empirical tests of independence using returns based on transaction data have generally found large and statistically significant negative correlation. Thus, it is reasonable to model returns as a process with random time intervals between transaction and serial correlation among returns on individual trades. Accordingly, an auto-regressive jump process that models common stock returns through time was proposed by Oldfield et al. (1977). This model consists of a diffusion process, which is continuous with probability 1 and jump processes, which are continuous with probability 1. The jump process is assumed to operate such that a jump occurs at each actual transaction, and allows the magnitudes of jumps to be auto-correlated. In addition, the model relies on the distribution of random time intervals between transactions. They suppose the dollar return of a common stock over a holding period of length s is the result of a process, which is a mixture process composed of a continuous and jump process,

$$\frac{dP}{P} = \alpha dt + \beta dW + z d\pi, \quad (8)$$

where P stands for share price, dW is the increment of a Wiener process with zero mean and unit variance, z is the percent change in share price resulting from a jump, $d\pi$ is a jump process (when $d\pi = 1$, a jump occurs; when $d\pi = 0$, no jump occurs) and $d\pi$ and dW are assumed to be independent. Jump amplitude is independent of $d\pi$ and dW , but jumps may be serially correlated, s is the elapsed time between observed price P_{t+s} and P_t . The number of jumps during the interval s is N , and $Z(i)$ are the jump size where $Z(0) = 1$ and $Z(i) \geq 0$ for $i = 1, \dots, N$. And the solution for Eq. 8 is

$$P(t+s) = P(t) \cdot Z(0)(1) \dots Z(N) \cdot \exp \left\{ (\alpha - \beta^2/2)s + \beta\sqrt{(s)}W \right\} \quad (9)$$

Divide Eq. 9 by $P(t)$ and take natural logarithms, then

$$\ln [P(t+s)/P(t)] = (\alpha - \beta^2/2)s + \beta\sqrt{(s)}W + \sum_{i=1}^N \ln Z(i). \quad (10)$$

According to the Eq. 10, we can see the third term of Eq. 10 is the jump process. If $N = 0$ then $\ln[P(t+s)/P(s)]$ is normally distributed with mean $(\alpha - \beta^2/2)s$ and variance β^2s . If the $\ln Z(i)$ are assumed to be identically distributed with mean μ and finite variance σ^2 , a general form of joint density for $\ln Z(i)$ can be represented by:

$$f(\ln Z(1), \dots, \ln Z(N)) = \int_{-\infty}^{\infty} f(\ln Z(1), \dots, \ln Z(N), W) dW,$$

with:

$$\begin{aligned} E[\ln Z(i)] &= \mu, \text{ for } i = 1, \dots, N, \\ \text{Var}[\ln Z(i)] &= \sigma^2, \text{ for } i = 1, \dots, N, \\ \text{Cov}[\ln Z(i), \ln Z(i-j)] &= \rho_j \sigma^2, \text{ for } j \geq 0. \end{aligned}$$

where ρ_j is the correlation between $\ln Z(i)$ and $\ln Z(i-j)$. The index i represents the jump number while the index j denotes the number of lags between jumps. The startling feature of this general joint density is the autocorrelation among jumps. Hence, some major conclusions are drawn from the data analysis: (1) A geometric Brownian motion process or a subordinated process does not alone describe the sample data very well. (2) Stock returns seem to follow an autoregressive jump process based on the sample means and variances of transaction returns. (3) In contrast to the previous empirical work which is not sufficiently detailed to determine the probability law for transaction returns, the probability density for the time intervals between jumps is gamma.

44.9 Jump Diffusion Models with Conditional Heteroscedasticity

44.9.1 Conditional Jump Dynamics

The basic jump model has been extended in a number of directions. A tractable alternative is to combine jumps with an ARCH/GARCH model in discrete time. It seems likely that the jump probability will change over time. Ho et al. (1996) formulate a continuous-time asset pricing model based on the work of Chamberlain (1988), but include jumps. Their work strongly suggested that both jump components and heteroscedastic Brownian motions are needed to model the asset returns. As the jump components are omitted, the estimated rate of convergence of volatility to its unconditional mean is significantly biased. Moreover, Chan and Maheu (2002) developed a new conditional jump model to study jump dynamics in stock market returns. They present a discrete-time jump model with time varying conditional jump intensity and jump size distribution. Besides, they combine the jump specification with a GARCH parameterization of volatility. Consider the following jump model for stock returns:

$$R_t = \mu + \sum_{i=1}^t \phi R_{t-i} + \sqrt{h_t} z_t + \sum_{k=1}^{n_t} Y_{t,k}, \quad (11)$$

$$z_t \sim NID(0, 1), Y_{t,k} \sim N(\theta_t, \delta^2).$$

Define the information set at time t to be the history of returns, $\Phi_t = \{R_t, \dots, R_1\}$. The conditional jump size $Y_{t,k}$, given Φ_{t-1} , is presumed to be independent and normally distributed with mean θ_t and variance δ^2 . Denote n_t as the discrete counting process governing the number of jumps that arrive between $t-1$ and t , which is distributed as a Poisson random variable with the parameter $\lambda_t > 0$ and density

$$P(n_t = j | \Phi_{t-1}) = \frac{\exp(-\lambda_t) \lambda_t^j}{j!}, j = 0, 1, 2, \dots \quad (12)$$

The mean and variance for the Poisson random variable are both λ_t , which is often called the jump intensity. The jump intensity is allowed time-varying. h_t is measurable with respect to the information set Φ_{t-1} and follows a GARCH(p,q) process,

$$h_t = w + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i},$$

where $\varepsilon_t = R_t - \mu - \sum_{i=1}^p \phi_i R_{t-i}$ contains the expected jump component and it affects future volatility through the GARCH variance factor. Moreover, based on a parsimonious ARMA structure, let λ_t be endogenous. Denote the following ARJI(r,s) model:

$$\lambda_t = \lambda_0 + \sum_{i=1}^r \rho_i \lambda_{t-i} + \sum_{i=1}^s \gamma_i \xi_{t-i}$$

$\lambda_t = E[n_t | \Phi_{t-1}]$ is the conditional expectation of the counting process. ξ_{t-i} represents the innovation to λ_{t-i} . The shock jump intensity residual is

$$\xi_{t-i} = E \left[n_{t-i} | \Phi_{t-i} \right] - \lambda_{t-i} = \sum_{j=0}^{\infty} j P(n_{t-i} = j | \phi_{t-i}) - \lambda_{t-i}. \quad (13)$$

The first term of Eq. 13 is average number of jumps at time $t - i$ based on time $t - i$ information. Therefore, ξ_{t-i} represents the unpredictable component about the conditional mean of the counting process n_{t-i} . Moreover, having observed R_t , let $f(R_t | n_t = j, \Phi_{t-1})$ denote the conditional density of returns given that j jumps occur and the information set Φ_{t-1} , we can get the ex post probability of the occurrence of j jumps at time t , with the filter defined as

$$P(n_t = j | \Phi_t) = \frac{f(R_t | n_t = j, \Phi_{t-1}) P(n_t = j | \Phi_{t-1})}{P(R_t | \Phi_{t-1})} \quad (14)$$

$$j = 0, 1, 2, \dots,$$

where, the definition of $P(n_t = j | \Phi_{t-1})$ is the same as Eq. 12. The filter in Eq. 14 is an important component of their model of time varying jump dynamics. Thus, the conditional density of return is

$$P(R_t | \Phi_{t-1}) = \sum_{j=0}^{\infty} f(R_t | n_t = j, \Phi_{t-1}) P(n_t = j | \Phi_{t-1}). \quad (15)$$

Equation 15 shows that this model is nothing more than a discrete mixture of distribution where the mixing is driven by a time varying Poisson distribution. Therefore, from the assumption of Eq. 11, the distribution of returns conditional on the most recent information set and j jumps is normally distributed as

$$f \left(R_t | n_t = j, \Phi_{t-1} \right) = \frac{1}{\sqrt{2\pi(h_t + j\delta_t^2)}} \times \exp \left(-\frac{\left(R_t - \mu - \sum_{i=1}^l \phi_i R_{t-i} - \theta_{tj} \right)^2}{2(h_t + j\delta_t^2)} \right).$$

Equation 14 includes an infinite sum over the possible number of jumps n_t . However, practically, they consider truncating the maximum number of jumps to a large value τ , and then they set the probability of τ or more jumps to 0. Hence, the first way to choose τ is to check Eq. 12 to be equal to 0 for $j \geq \tau$. The second check on the choice of τ is to investigate $\tau > \tau$ to make sure that the parameter estimate does not change.

The ARJI model illustrates that conditional jump intensity is time varying. Suppose that we observe $\xi_\tau > 0$ for several periods. This suggests that the jump intensity is temporarily trending away from its unconditional mean. On the other hand, this model effectively captures systematic changes in jump risk in the market. In addition, they find significant time variation in the conditional jump intensity and the jump size distribution in their application for daily stock market returns. Accordingly, the ARJI model can capture systematic changes, and also forecast increases (decreases) in jump risk into the future.

44.9.2 ARCH/GARCH Jump Diffusion Model

As described in Drost et al. (1998) there exists a major drawback of Merton's (1976) model which implies that returns are independent and identically distributed at all frequencies that conflict with the overwhelming evidence of conditional heteroscedasticity in returns at high frequencies, because all deviations from log normality of observed stock returns at any frequency can be attributed to the jumps in his model. Thus, several papers consider the size of jumps within the models that also involve the conditional heteroscedasticity.

Jorion (1988) considered a tractable specification combining both ARCH and jump processes for foreign exchange market:

$$\ln(P_t/P_{t-1})|t-1 = \mu + \sqrt{(b_t)}z + \sum_{i=1}^{n_t} \ln(Y_i),$$

$$b_t = E_{t-1}(\sigma_t^2) = \alpha_0 + \alpha_1(x_{t-1} - \mu)^2$$

in which α_1 is the autoregressive parameter inducing heteroskedasticity and the distribution of x_t is conditional on information at $t - 1$ and define x_t as the logarithm of price relative $\ln(P_t/P_{t-1})$. A jump size Y is assumed independently log normally distributed, $\ln Y \sim N(\theta, \delta^2)$, n_t is the actual number of jumps during the interval. z is a standard normal deviate. Consequently, his results reveal that exchange rate exhibit systematic discontinuities even after allowing for conditional heteroskedasticity in the diffusion process. In brief, in his work, the maximum likelihood estimation of a mixed-jump diffusion process indicates that ignoring the jump component in exchange rates can lead to serious mispricing errors for currency options. The same findings also can be found in Nieuwland et al. (1991) who allow for the model with conditional heteroscedasticity and jumps in exchange rate market. Also, an application of a GARCH jump mixture model has been given by Vlaar and Palm (1993). They point out that the GARCH specification cum normal innovation cannot fully explain the leptokurtic behavior for high-frequency financial data. Both the GARCH specification and the jump process can explain the leptokurtic behavior. Hence, they permit autocorrelation in the mean higher-order GARCH effect and Bernoulli jumps.

A weak GARCH model can be defined as a symmetric discrete-time process $\{y_{(h)t}, t \in h\mathbb{A}\}$ with finite fourth moment and with parameter $\zeta_h = (\phi_h, \alpha_h, \beta_h, \kappa_h)$, if there exists a covariance-stationary process $\{\sigma_{(h)t}, t \in h\mathbb{A}\}$ with

$$\sigma_{(h)t+h}^2 = \phi_h + \alpha_h y_{(h)t}^2 + \beta_h \sigma_{(h)t}^2, t \in h\mathbb{A}$$

and we denote $\kappa_h = \frac{E y_{(h)t}^4}{(E y_{(h)t}^2)^2}$ as the kurtosis of the process.

Roughly speaking, the class of continuous-time GARCH models can be divided into two groups. One is the GARCH diffusion in which the sample paths are smooth and the other, where the sample paths are erratic. Drost and Werker (1996) developed several properties of discrete-time data that are generated by underlying continuous-time processes that accommodate both conditional heteroscedasticity and jumps. Their model is as follows. Let $\{Y_t, t \geq 0\}$ be the GARCH jump diffusion with parameter vector $\zeta_h = (\phi_h, \alpha_h, \beta_h, \kappa_h)$ and suppose α_{h0} for some $h_0 > 0$. Then, there exists $\omega \in (0, \infty)$, $\theta \in (0, \infty)$, $\phi \in (0, \infty)$, $v \in (0, \infty)$ and c_h and K_h are given by

$$c_h = \frac{4\{\exp(-h\theta) - 1 + h\theta\} + 2h\theta\left\{1 + \frac{v + 2h\theta}{v\phi(2 + \phi)}\right\}}{1 - \exp(-2h\theta)},$$

$$\kappa_h = 3 + \frac{v}{h\theta} + 3v\phi(2 + \phi) \frac{\exp(-h\theta) - 1 + h\theta}{(h\theta)^2},$$

such that ζ_h (with $|\beta_h| < 1$) is determined by

$$\Psi_h = h\omega\{1 - \exp(-h\theta)\},$$

$$\alpha_h = \exp(-h\theta) - \beta_h, \frac{\beta_h}{1 + \beta_h^2} = \frac{c_h \exp(-h\theta) - 1}{c_h\{1 + \exp(-2h\theta)\} - 2},$$

where θ is the time unit and scale is denoted by ω . ϕ and v are slope parameters and ϕ will denote slopes in the (α_h, β_h) plane, while v determines the slope of the kurtosis at very high frequencies. Drost et al. (1998) employed the results of Drost and Werker (1996), which stated that for GARCH diffusion at an arbitrary frequency h , the five discrete-time GARCH parameters can be written in terms of only four continuous-time parameters, i.e., an over identifying restriction in GARCH diffusion, for proposing a test for the presence of jumps with conditional heteroscedasticity, which is based on the following Theorem 1.

Theorem 1 *Let $\{Y_t : t \geq 0\}$ be a continuous-time GARCH diffusion. Then $\theta > 0$ and $\lambda \in (0, 1)$ is defined by*

$$\begin{aligned}
\theta &= -\ln(\alpha + \beta) \\
\lambda &= 2\ln^2(\alpha + \beta) \left\{ \frac{\{1 - (\alpha + \beta)^2\}(1 - \beta)^2}{\alpha - \alpha\beta(\alpha + \beta)} \right. \\
&\quad \left. + 6\ln(\alpha + \beta) + 2\ln^2(\alpha + \beta) + 4(1 - \alpha - \beta) \right\} \\
\kappa &= 3 + 6 \frac{\lambda}{1 - \lambda} \frac{\exp(-\theta) - 1 + \theta}{\theta^2}
\end{aligned}$$

Thus, we set up the null and alternative hypotheses:

$H_0 : \{Y_t : t \geq 0\}$ is a GARCH diffusion model and $H_1 : \{Y_t : t \geq 0\}$ is a GARCH jump diffusion model.

From Theorem 1, by simple calculation, we yield the relation between functions K and κ :

$$K(\alpha, \beta, \kappa) = \kappa - \kappa(\alpha, \beta) = 0$$

for GARCH diffusion. Furthermore, in Drost and Werker (1996), they showed that $K(\alpha, \beta, \kappa)$ will be strictly larger than 0 for any GARCH jump diffusion model. As a result, H_0 is equivalent to $K(\alpha, \beta, \kappa) = 0$ and H_1 is equivalent to $K(\alpha, \beta, \kappa) > 0$. In other words, this test can be viewed as the kurtosis test for presence of jumps with conditional heteroscedasticity. As well, it indicates the presence of jumps in dollar exchange rate.

44.10 Other Jump Diffusion Models

As shown in Chacko and Viceria (2003), the jump diffusion process for stock price dynamics with asymmetric upward and downward jumps is

$$\begin{aligned}
\frac{dS_t}{S_t} &= \mu dt + \sigma dZ + [\exp(J_u) - 1]dN_u(\lambda_u) \\
&\quad + [\exp(-J_d) - 1]dN_d(\lambda_d).
\end{aligned}$$

$[\exp(J_u) - 1]dN_u(\lambda_u)$ and $[\exp(-J_d) - 1]dN_d(\lambda_d)$ represent a positive jump and a downward jump, respectively. $J_u, J_d > 0$ are stochastic jump magnitudes, which implies that the stock prices are non-negative, $\lambda_d, \lambda_u > 0$ are constant, and also determine jump frequencies. Furthermore, the densities of jump magnitudes,

$$f(J_u) = \frac{1}{\eta_u} \exp\left(-\frac{J_u}{\eta_u}\right) \text{ and}$$

and

$$f(J_d) = \frac{1}{\eta_d} \exp\left(-\frac{J_d}{\eta_d}\right)$$

are drawn from exponential distributions. Note that μ and σ are constants.

To estimate this process, they provide a simple, consistent procedure – spectral GMM by deriving the conditional characteristic function of that process.

References

- Ahn, C.M., and H.E. Thompson. 1988. Jump diffusion processes and the term structure of interest rates. *Journal of Finance* 43 (1): 155–174.
- Akgriray, V., and G.G. Booth. 1988. Mixed diffusion-jump process modeling of exchange rate movements. *Review of Economics and Statistics* 70 (4): 631–637.
- Bachelier, L. 1900. Theory of speculation. In *Random character of stock market prices*, ed. P. Costner. Cambridge: MIT Press, 1964, reprint.
- Ball, C.A., and W.N. Torous. 1983. A simplified jump process for common stock returns. *Journal of Financial and Quantitative Analysis* 18 (1): 53–65.
- . 1985. On jumps in common stock prices and their impact on call option pricing. *Journal of Finance* 40 (1): 155–173.
- Bandi, F.M., and T.H. Nguyen. 2003. On the functional estimation of jump-diffusion models. *Journal of Econometrics* 116: 293–328.
- Bates, D.S. 1991. The crash of '87: Was it expected? The evidence from options markets. *Journal of Finance* 46 (3): 1009–1044.
- . 1996. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. *Review of Financial Studies* 46 (3): 1009–1044.
- Chacko, G., and L.M. Viceria. 2003. Spectral GMM estimation of continuous time processes. *Journal of Econometrics* 116: 259–292.
- Chamberlain, G. 1988. Asset pricing in multiperiod securities markets. *Econometrica* 56: 1283–1300.
- Chan, W.H., and J.M. Maheu. 2002. Conditional jump dynamics in stock market returns. *Journal of Business and Economic Statistics* 20 (3): 377–389.
- Cox, J.C., J.E. Ingersoll Jr., and S.A. Ross. 1981. A re-examination of traditional hypotheses about the term structure of interest rate. *Journal of Finance* 36: 769–799.
- . 1985a. An intertemporal general equilibrium model of asset prices. *Econometrica* 53: 363–384.
- . 1985b. A theory of the term structure of interest rate. *Econometrica* 53: 385–407.
- Das, S.R. 2002. The surprise element: Jumps in interest rates. *Journal of Econometrics* 106: 27–65.
- Duffie, D., and R. Kan. 1996. A yield-factor model of interest rate. *Mathematical Finance* 6: 379–406.
- Duffie, D., J. Pan, and K. Singleton. 2000. Transform analysis and asset pricing for affine jump diffusion. *Econometrica* 68: 1343–1376.
- Drost, F.C., and B.J.M. Werker. 1996. Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics* 74: 31–57.
- Drost, F.C., T.E. Nijman, and B.J.M. Werker. 1998. Estimation and testing in models containing both jumps and conditional heteroscedasticity. *Journal of Business and Economic Statistics* 16 (2).
- Gourieroux, C., and J. Jasiak. 2001. *Financial econometrics*. 1st ed. Princeton: Princeton University Press.

- Ho, M.S., W.R.M. Perraudin, and S.E. Sorensen. 1996. A continuous time arbitrage pricing model. *Journal of Business and Economic Statistics* 14 (1): 31–42.
- Jarrow, R.A., and E.R. Rosenfeld. 1984. Jumps risks and the intertemporal capital asset pricing model. *Journal of Business* 57 (3): 337–351.
- Johannes, M. 2004. The statistical and economic role of jumps in continuous time interest rate model. *Journal of Finance* 59: 227–260.
- Johnson, T.C. 2002. Volatility, momentum, and time-varying skewness in foreign exchange returns. *Journal of Business and Economic Statistics* 20 (3): 390–411.
- Jorion, P. 1988. On jump processes in the foreign exchange and stock market. *Review of Financial Studies* 1 (4): 427–445.
- Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41 (5): 867–887.
- . 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3: 125–144.
- Niederhoffer, V., and M.F.M. Osborne. 1966. Market making and reversal on the stock exchange. *Journal of the American Statistical Association* 61: 897–916.
- Nieuwland, F.G.M.C., W.F.C. Verschoor, and C.C.P. Wolff. 1991. EMS exchange rates. *Journal of International Financial Markets Institutions and Money* 2: 21–42.
- Øksendal, B. 1998. *Stochastic differential equations*. 5th ed. New York: Springer.
- Oldfield, G.S., R.J. Rogalski, and R.A. Jarrow. 1977. An autoregressive jump process for common stock returns. *Journal of Financial Economics* 5: 389–418.
- Omberg, E. 1988. Efficient discrete time jump process models in option pricing. *Journal of Financial and Quantitative Analysis* 23 (2): 161–174.
- Osborne, M.F.M. 1959. Brownian motion in the stock market. *Operations Research* 7: 145–173.
- Press, S.J. 1967. A compound events model for security prices. *Journal of Business* 40 (3): 317–335.
- Singleton, K. 2001. Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102 (1): 111–141.
- Tucker, A.L., and L. Pond. 1988. The probability distribution of foreign exchanges: Tests of candidate processes. *Review of Economics and Statistics* 70: 638–647.
- Vasicek, O. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5: 177–188.
- Vlaar, P.J.G., and F.C. Palm. 1993. The Message in weekly exchange rates in the European monetary system: Mean reversion, conditional heteroscedasticity, and jumps. *Journal of Business and Economic Statistics* 11 (3): 351–360.



Daniel G. Weaver

Abstract

In the United States, the same stock can be traded at different locations. In the case of listed stocks, each location is a node in national network called the Intermarket Trading System (ITS). Unlisted stocks also trade at different nodes on the National Association of Securities Dealers Automated Quotation (NASDAQ) network. Each node of these two networks may have rules for breaking queuing ties among competing orders. Orders may be routed on the networks according to official rules (as with ITS) or order preferencing arrangements (both networks). This chapter examines the impact of priority rules on individual markets and networks. The development of the ITS and NASDAQ networks as well as the relevant literature is discussed. I conclude that network priority rules improve market quality if they result in consolidated markets.

Keywords

Consolidated · Fragmented · Intermarket Trading System · Market quality · NASDAQ · Networks · Nodes · NYSE · Preferencing · Priority rules

Assume that an investor wants to sell 100 shares of stock and a number of people are willing to buy it. Who should get to buy the 100 shares? If asked, the average person would say, the trader offering the highest price. What if there is more than one trader offering the same price? The average person would answer the trader who quoted the price first. However, many times, the trader quoting the best price first does not get to trade. An understanding of the determinants of trade sequencing (called priority rules) will assist investors in designing trading strategies. This chapter will review the different types of priority rules as well as the literature on the subject.

D. G. Weaver (✉)
Rutgers University, New Brunswick, NJ, USA
e-mail: daniel_weaver@business.rutgers.edu

Related to priority rules is the concept of order routing. The average person conceives of a market for stock as a single entity. While it is true that Microsoft is a The National Association of Securities Dealers Automated Quotation (NASDAQ)-listed stock, the NASDAQ system is only one node in a network, any one of which could execute a trade for Microsoft. Similarly, there are more than a handful of markets in the United States that trade New York Stock Exchange (NYSE)-listed GE. The markets for both Microsoft and GE can be thought of as networks with multiple nodes. Each node may or may not have similar priority rules to the other nodes in the network.

In addition, networks may have priority rules that govern the routing of orders within the network (as does the network for NYSE stocks) or may not have network-wide priority rules (this is the case on NASDAQ). In fact, recently, the U.S. Securities and Exchange Commission (SEC) and derivatively Congress have begun addressing the issue of whether networks should have priority rules. The SEC has proposed imposing network-wide priority rules on NASDAQ, while some market forces have tried to convince the Congress that not only should NASDAQ not have network-wide priority rules, they have also lobbied to eliminate NYSE-listed network priority rules.

From the brief discussion just presented, it is clear that a submitted order faces a maze of routing and priority rules. Therefore, this chapter will address these issues in hopes of shedding light on the relevant factors in designing an optimal network with regard to network routing and node priority rules.

In the following section, I shall discuss different types of priority rules in use in markets today as well as the literature on the subject. In Sect. 45.2, I will present a history of how the networks for NYSE-listed and NASDAQ stocks developed, which includes routing rules. In Sect. 45.3, I will discuss the current political and regulatory environment concerning stock networks. I will also discuss whether networks benefit from priority rules. In the final section, I conclude.

45.1 Priority Rules

Markets and network nodes use a variety of priority rules to match buyers with sellers.¹ Typically, price takes the highest priority: The buyer willing to pay the most is entitled to trade with the next seller willing to sell at the buyer's bid price and vice versa. However, if there is a tie in which more than one buyer is willing to buy at a different price, markets use a variety of different rules to decide who gets to purchase from the next seller. Here is a sample of the different secondary rules²:

¹See Domowitz (1993) for taxonomy of many of the different rules found in different markets.

²The following list is taken from an unpublished paper I co-authored with James Angel entitled "Priority Rules."

Time priority represents a first-come, first-served model. The first order submitted at a given price is the first one to be filled. The American Stock Exchange, Paris Bourse, Tokyo Stock Exchange, and Toronto Stock Exchange Computer-Assisted Trading System (CATS) prior to 1996 used some variation of this method. However, it is by no means clear that time priority is the most desirable secondary priority rule for a market. Indeed, few financial markets use pure time priority.³

Class priority gives priority to certain classes of traders over others. For example, on the Toronto Stock Exchange the Registered Trader has a higher priority than orders on the book in that he or she can participate in certain incoming trades up to half the minimum guaranteed fill. On the NYSE, however, the specialist cannot trade ahead of the limit order book. In a dealer market such as the old NASDAQ system (prior to the new order handling rules), dealers could take priority over customer orders – even if customer orders at a better price – because the customers have no means of bypassing the intermediaries. We will discuss dealer priority in more depth later in the paper.

Random priority randomly assigns an order among the traders willing to trade at a given price. Each floor trader willing to trade at a given price has an equal probability of filling the next order. This is effectively what happens in the “open outcry” method found in floor-based futures trading pits such as the Chicago Board of Trade.

Sharing or pro rata priority is also a common practice on many trading floors including the Stock Exchange of Hong Kong and the old Toronto Stock Exchange floor-based system. A sharing priority rule could allocate equal shares to each order on the book. Alternatively, the allocation could be proportional to the total size of a member’s orders on the book (pro rata sharing.) However, even if a trading floor has a time priority rule, it may be virtually impossible to determine who was first. For example, a large order may arrive at a trading venue where there are several traders willing to fill the order. Therefore, a large order may be de facto shared among many traders.

Size priority grants priority to orders based on their size. Priority could be granted to the largest order, which has the advantage of giving traders an incentive to place larger orders.⁴ Alternatively, priority could be granted to the order that matches the incoming order in size. This minimizes the number of trade tickets to be processed. A variation of this secondary priority rule is used on the New York Stock Exchange.

Exposure priority grants priority to orders that are revealed to other market makers and reduces the priority for those traders who want to hide their orders. On the old

³For example, on the old TSE CATS system, time priority did not expire, while on the Amex time priority lasts only until the next trade.

⁴Note that size priority is different than pro rata sharing in that an entire incoming order may go to a single trader as opposed to be shared.

CATS system and the Paris Bourse, traders can hide a portion of their orders from exposure on the electronic systems.

These different secondary priority rules have strong implications for the ways that investors compete to obtain an order fill. In a pure time priority market, an investor who is the first to put in a bid at a higher price is first in line to fill the next market sell order. With random or sharing priority there is much less incentive to pay up by bidding higher. This is because there is a positive probability that a trader can obtain a fill, within the same time to execution as bidding higher, by merely matching existing quotes.

Therefore, it can be seen that priority rules can have an impact on market quality. For example, in systems with time as the secondary priority rule, traders have incentives to improve on the price since merely matching a current best price puts them at the back of the queue. This could lead to narrower spreads. Similarly, in systems where public orders take priority over market maker or specialist orders, there will be more public orders submitted. This can lead to more liquidity being supplied.

45.2 Literature on Priority Rules

Cohen et al. (1985) find support for the notion that time priority leads to more price competition and, hence, narrower spreads. They use a simulated queuing model to show that systems that do not enforce time priority have wider spreads relative to those that enforce time priority.

Angel and Weaver (1998) and Panchapagesan (1998) compare market quality and investor behavior differences between systems that use time priority as their secondary rule with systems that use pro rata sharing. In particular, Angel and Weaver examine the 29 July 1996 switch from time to sharing priority for stocks in the Toronto Stock Exchange's (TSX) CATS. Panchapagesan also examines the TSX but compares a matched sample of CATS stocks with stocks traded on the TSX floor. During Panchapagesan's sample period the TSX floor used sharing priority rules while CATS used time priority.

Both studies find that a sharing priority rule results in less price competition compared with a time priority rule. They also find that a sharing priority rule results in more gaming behavior by investors in an attempt to get their orders filled. For example, a sharing priority rule encourages investors to submit larger orders and then cancel them when their desired volume is filled. Panchapagesan (1998) additionally concludes that the lack of price competition in sharing priority rule systems results in wider bid ask spreads than under time priority.

Cordella and Foucault (1999) develop a theoretical model of dealer competition and also conclude that spreads will be wider under a random allocation rule than under a price/time priority system. The intuition is that under a random allocation rule dealers can always match other dealers' quotes without losing priority.

Harris (1994) addresses the relationship between priority rules, tick size, and depth.⁵ Harris points out that large ticks and time priority protect traders that place limit orders. If a trader wants to trade ahead of another in a time priority rule system, he or she must improve on the price. A large tick makes obtaining precedence costly.⁶ Harris then argues that time priority encourages traders to quote more size, which leads to greater quoted depth.

Two points are evident from the above discussion. First, traders will change their behavior as rules change. Second, the behavior of these traders impacts market quality and hence the terms of trade for unsophisticated traders.

45.3 Networks

Thus far, the discussion of priority rules has assumed that there is a single market for stocks. That is markets are consolidated. What if there are multiple markets? In this section, I will discuss networks of markets and how priority rules may apply to them. I will also consider the development of two major network structures: one for listed stocks such as those on the NYSE and another for over-the-counter (OTC) stock.

45.3.1 The Network for Listed Stocks

During the first half of the twentieth century, the role of regional stock exchanges changed and their number decreased dramatically. When regional stock exchanges like the Boston Stock Exchange were first established, there was poor telecommunications in the United States and travel was expensive. As a result, it was very difficult for investors away from a company's headquarters to find out anything about the company. Therefore, regional stock exchanges were established as a place to trade local companies. As telecommunications improved and travel became less expensive, it became easier to find out about companies located in distant geographic locations. As companies grew, they switched listed on the American or New York Stock Exchanges where they could obtain the prestige of a national listing.

So, the regional exchanges experienced a dramatic drop in listings. Perhaps to provide another source of revenue, local traders began trading NYSE-listed stocks. Multiple trading venues for the same NYSE-listed stock led to frequent differences in prices across markets. Stories are abound of traders paying for open phone lines between the NYSE and one of the regional stock exchanges so that they could capitalize on the discrepancies. Over time, Congress observed that prices for

⁵Tick size is the minimum price increment.

⁶Assume a \$0.125 tick size. Then in order to step ahead of an existing buy order, a trader must be willing to pay \$12.50 more for each 100 shares he obtains. If the tick is only \$0.01 then that same trader must only pay \$1.00 more for each 100 shares he obtains.

NYSE-and Amex-listed stocks varied widely across the exchanges that traded them. This led to the passage of the Securities Act Amendments of 1975 in which Congress ordered the SEC to create a National Market System (NMS) that, in part, would allow investors to execute trades on markets that displayed the best price.

After deliberation, on 26 January 1978, the SEC issued the Exchange Act Release No. 14416 that required markets to create a network that would “permit orders for the purchase and sale of multiple-traded securities to be sent directly from any qualified market to another such market promptly and efficiently.”⁷ Two months later, the American, Boston, NYSE, Pacific, and Philadelphia Stock Exchanges submitted a “Plan for the Purpose of Creating and Operating an Intermarket Communications Linkage.” This became known as the Intermarket Trading System (ITS).

The ITS allowed exchanges to route orders to each other. It was in effect an e-mail system in which the specialist on an exchange could ask a specialist on another exchange if they would be willing to trade at their quoted price.⁸ A few months later, the SEC also created the Consolidated Quote System (CQS), which collected the best quoted prices to sell (called the offer) and buy (called the bid) securities. The CQS then constructed the best bid and offer (BBO) and disseminated it to the exchanges and data vendors who disseminated it to the public.

Although the ITS established a network and a method for routing orders, there initially was no rule indicating under what circumstances participants were to route orders to another exchange. Nor was the OTC market a part of the linkage. The NASDAQ traded many NYSE and Amex stocks. After the passage of the Securities Act Amendments of 1975, they began work on a Computerized Automated Execution System (CAES) to interface with ITS. Finally, on 28 April 1981, the SEC issued an order requiring that CAES become a part of ITS.

That same month, the SEC issued Securities Exchange Act Release No. 17703 that prohibited ITS/CAES participants from executing orders at prices inferior to those displayed on another network node. This requirement became known as the trade through rule and established a network-wide priority system for investor orders. The rule requires an exchange to either match a better price or route the order to the exchange displaying the better price.

The effect of the trade through rule was and is to establish price as the first priority rule across the ITS network. In turn, each regional exchange (a node) has its own priority rules.⁹ So, an order reaching any node of the ITS network was first subject to a network-wide price priority rule and then the node’s priority rules.

⁷Exchange Act Release No. 14416 at 4358.

⁸A specialist is the designated primary dealer on a stock exchange. They have complete knowledge of all investor orders and generally have obligations to maintain orderly markets.

⁹NASDAQ will be discussed in the next section.

45.3.2 The Network for OTC Stocks

In the previous section, I discussed how a network for exchange-listed stocks developed. While exchanges list many stocks in the United States, the majority of stock issued are traded OTC, so-called because you went to your broker's office and purchased them at the front counter. In the early part of the twentieth century, there was no organized way to buy and sell OTC stocks and so many times investors or brokers resorted to newspaper ads to accomplish the task.

In 1913, the National Quotation Bureau was formed by two businessmen who collected and published daily quotations from dealers of securities in five different cities. Their publication became known as the Pink Sheets because of the color of the paper they were printed on. For each stock the Pink Sheets listed the brokers trading the stock and representative quotes. The quotes were old, but at least brokers had a list of other brokers who were interested in trading a particular stock issue. Brokers would contact those listed in the Pink Sheets to get current quotes. Because of the large number of listed brokers for some stocks and the amount of time necessary to call each one, a rule was developed over time that required brokers to contact at least three (but not all) of the brokers listed on the Pink Sheets in order to try and find the best price for customers. Due to the lack of continuous investor interest in OTC stocks, the market developed as a dealer market who would act as intermediaries between investors – buying and selling stocks to earn a profit.

The vast majority of stock issues (but not stock volume) were traded this way for almost 60 years. As companies grew, they typically listed on an exchange. Over time, though telecommunications improved and the Pink Sheets expanded their coverage to nationwide. This increased the number of dealers making markets in a particular stock and made the goal of finding investors the best price more difficult. So, in 1971, the National Association of Securities Dealers (NASD) created an automated quotation system with the acronym NASDAQ for trading the more active OTC issues. The system allowed dealer members to input contemporaneous quotations for stocks they made a market in. NASDAQ was similar in its aggregating function to the CQS for listed stocks. Brokers still needed to telephone dealers to trade. And NASDAQ was a dealer market. While the exchanges established prices based on a combination of public limit orders and specialist quotes, NASDAQ displayed *dealer* quotes.^{10, 11}

Initially, NASDAQ dealers could ignore customer limit orders. Customers learned that limit orders were not executed and did not submit them. In 1994, an

¹⁰Limit orders are orders to buy or sell a security at a specified price or better. Public customers submit limit orders, specialists and NASDAQ market makers submit quotes.

¹¹The quotes are predominately based on public orders. As evidence of this, consider that during 2003, NYSE specialists were involved in less than 20% of all trades.

investor sued his broker and as a result NASDAQ established a rule which came to be known as Manning I. The rule prevented NASDAQ dealers from trading through *their* customer limit orders at better prices – much like ITS trade through rules do.¹² However, after the passage of Manning I, NASDAQ dealers could still trade at the same price as customer limit orders they held, i.e., there was no public order priority rule. This was in contrast to the exchanges that had public priority rules. NASDAQ customers were still reluctant to submit limit orders. A year later, another rule, Manning II, gave public limit orders priority, but only within a dealer firm. In other words, a customer submitting a limit order to Dealer X could still see trades occurring at other dealers at the same price as the customer's limit order. Thus, Manning II still discouraged public limit order submission and as a result they were not a major supplier of liquidity on the NASDAQ market.

So it can be seen that although NASDAQ was a network of dealers, it had no market-wide priority rules as did the exchanges. In addition, proprietary trading systems were established that allowed NASDAQ dealers to trade between themselves at prices that were better than the best quotes on NASDAQ. Like it had for exchanges, Congress and the SEC acted and established a method for investors to access the best quoted prices. However, the landscape for NASDAQ stocks in the 1990s was different than that for exchange listed stocks in 1975. While a number of exchanges were trading the same stocks in 1975, there were really only two players for NASDAQ stocks by the mid-1990s. Other than NASDAQ itself, where dealers traded with the public, the only other place NASDAQ stocks were traded was on a proprietary system called Instinet.¹³

Most, if not all, NASDAQ dealers also were already connected to Instinet; so, rather than create a new network, the SEC required that Instinet quotes be made part of the BBO for NASDAQ stocks. The Order Handling Rules (OHR), enacted in early 1997, also required NASDAQ dealers to expose customer limit orders to the public by including them in their quotes. Rather than specifically including Instinet into the calculation of the BBO, the SEC generalized the rule to include *any* system for displaying limit orders for NASDAQ stocks. These systems were referred to as Electronic Communications Networks (ECNs) and a number of new systems were established in anticipation of the OHR or shortly after its passage.

These ECNs unleashed the potential of public limit orders. After the OHR, spreads dropped dramatically with most of the drop attributed to public limit orders competing with dealer quotes.¹⁴ ECNs grew in market share from around 20% in 1997 to 80% today. The OHR created a much larger network of systems than ITS

¹²The NASDAQ trade, through rule, only applied to an individual broker. That is, Broker X was not allowed to trade through any customer of Broker X, but was not prevented from trading through customer limit orders held by Broker Y.

¹³The Midwest Stock Exchange traded some NASDAQ stocks, but was a distant third in market share.

¹⁴See Barclay et al. (1998).

ever faced. At the time of the OHR, the SEC did not require a trade through rule for the NASDAQ/ECNs network as it did for ITS.

45.3.3 Do Networks Need Priority Rules?

During 2004 the SEC proposed Regulation NMS. Part of the rule proposes to extend the ITS trade through rule to NASDAQ. On 10 February 2004, Congressman Richard Baker sent SEC Chairman William Donaldson a letter calling the ITS trade through rule “antiquated” and calling for its complete repeal. Congressman Baker suggested that execution speed was just as important as price and that investors should be allowed to choose whether they wanted price or speed to be the primary routing rule.

There is support for Congressman Baker’s position in the academic literature. For example Hatch et al. (2001) compare trade executions for NYSE-listed stocks between different nodes on the ITS/CAES network. They find that investors receive better prices on the NYSE, but the regional exchanges offer more speed of execution and larger execution sizes. In addition, Battalio et al. (2002) examined limit orders execution. They find that at-the-inside limit orders do better on regional exchanges in terms of speed of execution (perhaps due to shorter queues) than on NYSE, but quote improving limit orders do better on the NYSE where they execute faster and more profitably.

Congressman Baker’s letter then raises the question: Do networks need priority rules? On its face, it would seem obvious that investors should be allowed to send orders wherever they choose. However, overall market quality must be balanced against the needs of individual traders. If the needs of the individual do not cause harm to the overall population then the individual should be allowed to route orders as they wish. If however, the overall population of traders is harmed by the choices of the individual then the needs of the majority out-weigh the individual’s needs.

A similar argument is used to justify nonsmoking areas. While an individual aware of smoking’s risks has the right to smoke, the impact of second-hand smoke on nonsmokers is such that nonsmokers will be harmed if smokers exercise their right to smoke around them. Therefore, various laws have been enacted to protect non-smokers from the harmful affects of second-hand smoke. The greater good comes down on the side of providing nonsmokers with a smoke-free environment.

Following the smoking analogy, the ability of traders to choose their priority rules should be weighed against overall market quality. For this purpose, the literature on consolidation and fragmentation becomes useful. In a consolidated market, order flow is concentrated in a single location. In a fragmented market, order flow is split up between multiple locations. The number of choices between consolidated and fragmented is a continuum not a bifurcation.

A number of papers have been written on the subject of fragmented versus consolidated markets. They generally conclude that consolidated markets offer better market quality than fragmented markets. For example, Madhavan (2000) developed a theoretical model that shows that fragmented markets have higher

volatility than consolidated markets. Wei and Bennett (2003) find empirical support for Madhavan's conjecture. In particular, they find that stocks that switch from the fragmented NASDAQ to the comparatively consolidated NYSE, experience a reduction in spread and volatility.¹⁵ Barclay et al. (2003) examined stock price and volume around quarterly expirations of the S&P 500 futures contract (so called witching days). They found that NYSE prices are more efficient than NASDAQ prices. They attributed the superior performance of the NYSE to the larger degree of order flow consolidation found there relative to NASDAQ.

Battalio et al. (1998) examined Merrill Lynch's decision to route all orders for NYSE-listed stocks to the NYSE rather than to a regional exchange where they could effectively internalize the order flow. They found that, consistent with other studies, the NYSE routing decision resulted in investors obtaining better prices and spreads narrowing.

Murphy and Weaver (2003) examined the TSX rule that require brokers receiving market orders of 5,000 shares or less, to either improve on price or send the order to the TSX for execution against limit orders. Following the adoption of the rule, the affected stocks experienced an immediate increase in depth and reduction in spread.¹⁶ In addition to the TSX, many other exchanges around the world have so-called concentration rules.

Therefore, the extant literature suggests that overall market quality is higher in consolidated versus fragmented markets. The NYSE market share of its listed stocks is around 80% and they display the best price over 90% of the time. Therefore the market for NYSE stocks can be considered relatively consolidated. Although never empirically tested, there appears to then be a link between percentage of the time a market displays the best price and its market share. So the ITS network price priority rule may be the mechanism that causes the consolidation of NYSE stocks. If orders are routed away from the NYSE to another exchange or market maker then the market for NYSE stocks will become more fragmented. The academic literature suggests that an increase in fragmentation will result in wider spreads and higher volatility. It has been shown, time and time again, that investors factor execution costs into their required cost of supplying funds to firms.¹⁷ Therefore, higher execution costs will translate into higher costs of capital for firms and stock prices will fall.

¹⁵They use a Herfindahl–Hirschman Index as well as the number of nodes trading a stock to measure the degree of fragmentation before and after the switch to the NYSE and find that the gains in spread width and volatility are greater for firms experiencing more fragmentation prior to the decision to list on the NYSE.

¹⁶The findings of Murphy and Weaver also suggest that TSX members eventually began using order routing technology that allowed them to capitalize on the TSX crossing priority rule. This action dampened the impact of consolidation and spreads widened again.

¹⁷See Amihud (2002), Amihud and Mendelson (1986) and Amihud et al. (1997), among others.

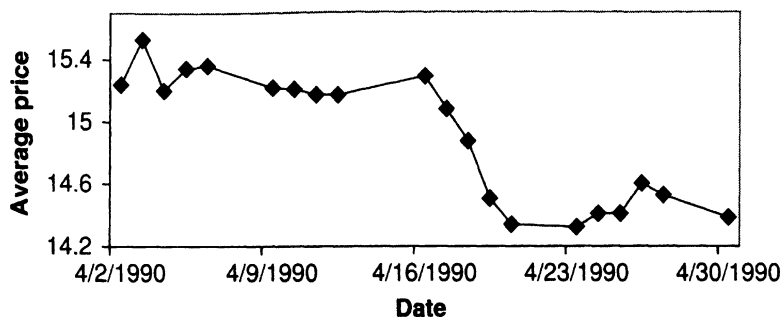


Fig. 1 Average daily prices of stocks in our sample for April 1990

Figure 1 illustrates the relationship between execution costs and stock prices.¹⁸ On 11 April 1990, the TSX enacted rules that resulted in effective execution costs rising by about 0.25% points. Within a week, prices declined by over 6%.

It can, therefore, be concluded that a network priority rule based on price results in improved market quality. Although a direct empirical link can only be proven by examining what happens if the network price priority rule is removed, logical inferences can be drawn from examining the behavior of those traders that supply liquidity to the market. The following section discusses the behavior of limit order traders who are the major supplier of liquidity on the NYSE.¹⁹

45.3.4 Liquidity Supplied by Limit Order Traders

There is an old adage that “liquidity begets liquidity.” In other words, limit order traders will submit limit orders where market orders exist. It is similar to the fact that the more traffic exists on a highway, the more gas stations will exist. If the traffic goes away so will the gas stations. Similarly, if market orders get routed away from the venue with the best price, limit orders will leave that venue as well. There will be less price competition and, as a result, spreads will widen.

Limit orders are shock absorbers for liquidity events. Without limit orders to absorb trades from liquidity demanders, large orders will increasingly push prices away from current prices.²⁰ While it may be argued that price impact is a fact of life

¹⁸Taken from Madhavan et al. (2005).

¹⁹Recall that NYSE specialists are involved in less than 20% of all trades.

²⁰Assume that there are 100 shares offered at \$19,200 at \$19.05, 100 at \$19.10, and 300 at \$19.15. A market order to buy 500 shares will take out the sell orders from \$19 to \$19.15, leaving the best offer at \$19.15 until new offers to sell arrive. This is sometimes referred to as walking the book.

for institutions, small traders who submit order in the same direction, but just behind the large order may suffer financial loss. The small order will execute at an inferior price before sufficient liquidity can be sent to the market by traders. It can then be seen that thin markets are more susceptible to liquidity event volatility than deeper markets.²¹ Thus, markets with more depth are desirable.

The TSX market concentration rules best illustrate the above points. Prior to the adoption of the rule, it was common practice for member firms to execute market orders and marketable limit orders from the member's own inventory (called internalization). Limit order traders realized that even if they had the best quoted price, many orders would never reach them and they would not get timely executions.²² The TSX adoption of its concentration rules caused more market and marketable limit orders to be submitted to the exchange where they could execute against limit orders. The increase in order flow to the exchange caused more limit order traders to compete for the order flow. This, in turn, resulted in narrower spreads and more quoted depth.

This section of the paper suggests that networks without priority rules discourage limit order submission which results in higher effective execution costs for the average investor. A few large players may benefit from the absence of a network priority rule, but it will be at the expense of the majority of long-term investors. Therefore, it can be seen that overall market quality benefits from network priority rules.

45.3.5 A Final Note on the Need for Speed

In the current drive to eliminate priority rules for the ITS network, the most common reason cited is a desire to get a trade done quickly – perhaps in a second or less. Is this advantageous? Perhaps examining a graph of a random stock on a random day would help. Figure 2 is the graph representing all trades in Juniper Inc. (JNPR) for 3 February 2003 from 10:00 AM until 10:01:30 AM.

It can be seen that getting an order filled at 10:00:51 compared to 10:00:52 may save you \$0.02 on that trade. However, if we examine JNPR over the entire day it can be seen that prices fluctuated by \$0.20 over the day, a factor of 10. So, price changes over small-time increments are much smaller than over longer increments (Figure 3). In that case, what type of trader benefits from small price changes and, hence, needs speed? The answer is arbitrageurs and hedge funds. As mentioned earlier, if we allow orders to be routed for other than best price, then limit order traders will reduce the amount of liquidity they supply, increasing execution costs. It can then be seen that this “need for speed” benefits the few at the expense of the many

²¹ Assume a deeper market of 600 shares offered at \$19. Then a 500 share order will not move the price.

²² Recall that a similar situation existed on NASDAQ before the adoption of the OHR.

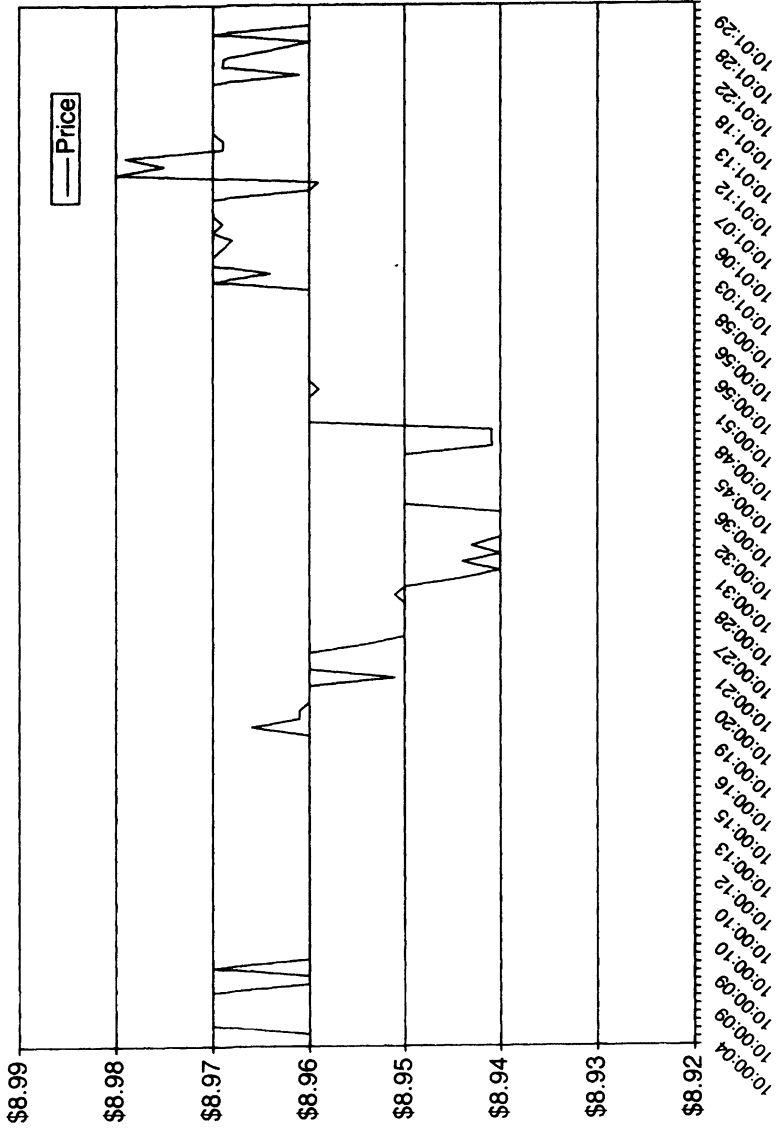


Fig. 2 JNPR 3 February 2003 10:00 to 10:01:30 AM

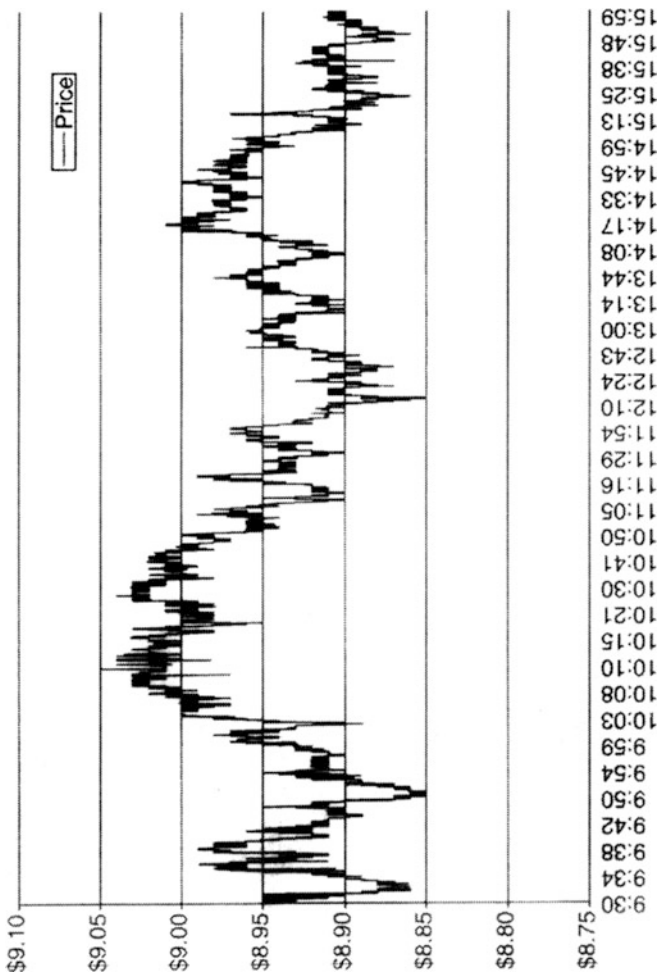


Fig. 3 Juniper corporation 3 February 2003

45.4 Conclusion

In this chapter, I have considered priority rules as they apply to individual markets (nodes) and networks. The literature on priority rules suggests that the adoption of some priority rules can improve market quality within a node. After a discussion of the development of the ITS and NASDAQ networks, I consider whether network priority rules matter. Since a network's priority rules (or routing rules) can result in a concentration of orders, I discuss the literature on consolidated versus fragmented markets. I conclude that network priority rules improve the market quality if they result in consolidated markets. This further suggests that the current price priority rule on ITS should, therefore, be retained and extended to the NASDAQ network as proposed by the SEC.

References

- Amihud, Y. 2002. Illiquidity and stock returns: Cross-sectional and time-series effects. *Journal of Financial Markets* 5: 31–56.
- Amihud, Y., and H. Mendelson. 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17: 223–250.
- Amihud, Y., H. Mendelson, and B. Lauterbach. 1997. Market microstructure and securities values: evidence from the Tel Aviv Stock Exchange. *Journal of Financial Economics* 45: 365–390.
- Angel, J., and Weaver, D. 1998. Priority Rules! Working Paper, Rutgers University.
- Barclay, Michael J., William G. Christie, Jeffrey H. Harris, Eugene Kandel, and Paul H. Schultz. 1999. The effects of market reform on the trading costs and depths of Nasdaq Stocks. *Journal of Finance* 54: 1–34.
- Barclay, M.J., Hendershott, T., and Jones, C.M. 2003. Which witches better? A cross-market comparison of extreme liquidity shocks. Working Paper, University of Rochester.
- Battalio, R., J. Greene, and R. Jennings. 1998. Order flow distribution, bid-ask spreads, and liquidity costs: Merrill Lynch's decision to cease routinely routing orders to Regional Stock Exchanges. *Journal of Financial Intermediation* 7: 338–358.
- Battalio, R., J. Greene, B. Hatch, and R. Jennings. 2002. Does the limit order routing decision matter? *Review of Financial Studies* 15 (1): 159–194.
- Cohen, K., R. Conroy, and S. Maier. 1985. Order flow and the quality of the market. In *Market making and the changing structure of the securities industry*, ed. Ho Amihud and Schwartz. Lexington: Lexington Books.
- Cordella, T., and T. Foucault. 1999. Minimum price variations, time priority and quote dynamics. *Journal of Financial Intermediation* 8 (3): 141–173.
- Domowitz, I. 1993. A taxonomy of automated trade execution systems. *Journal of International Money and Finance* 12: 607–631.
- Harris, L. 1994. Minimum price variations, discrete bid-ask spreads, and quotation sizes. *Review of Financial Studies* 7: 149–178.
- Hatch, B., R. Battalio, and R. Jennings. 2001. Postreform market execution quality: Multi-dimensional comparisons across market centers. *The Financial Review* 36 (3): 123–151.
- Madhavan, A. 2000. Consolidation, fragmentation, and the disclosure of trading information. *Review of Financial Studies* 8 (3).

- Madhavan, Ananth, David Porter, and Daniel Weaver. 2005. Should securities markets be transparent? *Journal of Financial Markets* 8: 265–287.
- Murphy, A., and Weaver, D. 2003. Order flow consolidation and market quality: An empirical investigation. Working Paper, Manhattan College.
- Panchapagesan, V. 1998. What if time priority is not enforced among traders? Working Paper, Washington University.
- Wei, L., and Bennett, P. 2003. New York stock market structure, fragmentation and market quality – Evidence from recent listing switches. Working Paper, New York Stock Exchange.

The Momentum Trading Strategy

46

K. C. John Wei and Linti Zhang

Abstract

A cross-sectional momentum strategy that buys past winners and simultaneously sells past losers based on stock performance in the past 3–12 months is profitable in the USA and many international markets. The profitability of the cross-sectional momentum strategy varies across firms, countries, and market states. Recent studies find that a time-series momentum strategy that longs securities with positive excess returns in the previous 12 months and shorts securities otherwise also generates persistent and high profits. This survey entry reviews the literature on the momentum strategies and the possible explanations for the momentum profitability.

Keywords

Past winners · Past losers · Momentum strategy · Individual momentum · Industrial momentum · Time-series momentum · Intraday momentum · International momentum · Underreaction · Overreaction · Overconfidence · Self-attribution · Valuation Uncertainty · Conservatism · Representative heuristic · Gradual information diffusion · Individualism

46.1 Introduction

“Trend is your friend” is a very popular saying in Wall Street since the inception of stock markets. However, whether this momentum trading strategy that buys past winners and sells past losers is really profitable was controversial until recently. Jegadeesh and Titman (1993) were the first to comprehensively examine the

K. C. J. Wei (✉) · L. Zhang

School of Accounting and Finance, The Hong Kong Polytechnic University, Kowloon, Hong Kong
e-mail: john.wei@polyu.edu.hk; johnwei@ust.hk; lintzhang@polyu.edu.hk;
linti1002.zhang@polyu.edu.hk

profitability of the cross-sectional momentum trading strategy based on the past 3- to 12-month performance. They document that momentum strategies implemented in the US market from 1965 to 1989 generated a positive profit of about 1% per month over 3- to 12-month holding periods. In their follow-up study, Jegadeesh and Titman (2001) find that momentum strategies continued to be profitable after 1990 with past winners outperforming past losers by about the same magnitude as in the earlier period.

Rouwenhorst (1998) studies individual stock momentum with a sample of stocks listed on 12 European exchanges during the period from 1978 to 1995. The results demonstrate that momentum profits of about 1% per month are not limited to a particular market, but instead they are present in all 12 markets in the sample. Rouwenhorst (1999) also finds that momentum strategies are profitable though not to the same degree in 20 emerging markets. Chui, Titman, and Wei (2002) examine the profitability of momentum strategies in eight Asian countries: Hong Kong, Indonesia, Japan, Korea, Malaysia, Singapore, Taiwan, and Thailand. Their evidence indicates that the momentum effect is present in all of the Asian countries except for Korea and Indonesia, but it is generally weak and is statistically significant only for Hong Kong, Malaysia, Singapore, and Thailand in the pre-Asian financial crisis period. Interestingly, they find that the common law/civil law distinction provides an indicator of whether or not a market exhibited a momentum effect.

Moskowitz and Grinblatt (1999) find that industry momentum strategies, which advocate buying stocks from past winning industries and selling stocks from past losing industries, appear to be highly profitable. This industry momentum accounts for much of the profitability of individual stock momentum strategies in the USA. Once returns are adjusted for industry effects, momentum profits from individual equities are significantly weaker and, for the most part, are statistically insignificant. However, Grundy and Martin (2001) have a different view on the contribution of industries to individual momentum profits. They argue that a 1-month interval between the ranking period and the holding period has a pivotal role in the conclusion that industry momentum strategies are profitable. Industry momentum strategies are significantly profitable only when the ranking period is contiguous to the holding period as documented by Moskowitz and Grinblatt (1999). However, with a 1-month gap between the two periods, industry momentum strategies no longer earn significant profits. Grundy and Martin (2001) conclude that industry effects are not the primary cause of the individual momentum profitability. Liu and Wei (2004) document that industries in 12 European markets, like their counterparts in the US market, also explain the profitability of individual momentum strategies. Specifically, past winner industries outperform past loser industries by more than 1% per month. However, unlike their counterparts in the US market, industries cannot solely explain the profitability of individual momentum strategies in 12 European markets. In addition, industry momentum strategies can still earn significant profits even with a 1-month interval between the formation and holding periods. Nijman, Swinkels, and Verbeek (2004) examine momentum profits in 14 European countries. They also find that individual momentum is not subsumed by industry or country momentum. Lewellen (2002) extends the momentum strategies to size and/or book-to-market

(BM) sorted portfolios. He documents strong size and BM momentums distinct from industry momentum in that neither can subsume the other. Because momentum also exists for diversified portfolios, he argues that the momentum effect cannot be entirely explained by firm- or industry-specific returns. Asness, Liew, and Stevens (1996), Richards (1997), and Chan, Hameed, and Tong (2000) document that momentum strategies are profitable when implemented on stock market indices.

Extant literature has been focusing on cross-sectional momentum (XSMOM) strategies that trade assets based on their past performances relative to those of others. More recently, several studies examine time-series momentum (TSMOM) strategies that trade assets based on their own past returns. Moskowitz, Ooi, and Pedersen (2012) (MOP, hereafter) document significant TSMOM in futures prices of equity indexes, commodities, currencies, and government bonds during the sample period from 1965 to 2009. Specifically, in a pool regression to use (volatility-scaled) lagged returns to predict (volatility-scaled) subsequent returns of the 58 most liquid futures contracts, the slope coefficient on lagged returns is statistically significantly positive up to 12 months. For each of the financial instrument considered, a time-series momentum trading strategy that longs (shorts) a security in case of positive (negative) historical excess returns generates significantly positive alphas relative to well-known asset-pricing models. The result is robust to various lookback and holding periods. The abnormal returns do not seem to be compensation for crash or tail risk, as the TSMOM strategy is profitable during both market downturns and upturns, and is not related to investor sentiments. They also find that auto-covariance in security returns make large contributions to both XSMOM and TSMOM, suggesting that the two types of momentums are related. However, the XSMOM factor cannot entirely explain the significantly positive abnormal return of the TSMOM strategy; while after adjusting for the TSMOM factor, alphas of the XSMOM strategy shrink and even become insignificant in some markets. They conclude that the TSMOM may be a distinct feature to describe asset-pricing behavior. Georgopoulou and Wang (2017) also document significant time-series momentum effects in equity and commodity indexes of 45 developed and emerging markets. Lim, Wang, and Yao (2018) implement the TSMOM strategy on US common stocks and document strong time-series momentum at the firm level.

However, several follow-up studies challenge the profitability of the TSMOM strategy. For instance, Kim, Tse, and Wald (2016) argue that MOP's results (2012) are driven by their volatility scaling. Using equal weighting instead, they find that the TSMOM strategy earns alphas similar to buy-and-hold alphas. Huang, Li, Wang, and Zhou (2020) use the same data set of MOP (2012) and reexamine the statistical and economic significances of the TSMOM strategy. They find that the predictability of lagged returns for future returns is not statistically significant at the individual asset level. The *t*-statistic of the coefficient on lagged returns in the pooled regression across assets is lower than the critical value under a bootstrap test. Besides, a strategy based on historical average returns, which does not require predictability, performs as well as the equal-weighted TSMOM strategy. Overall, the findings reveal weak evidence of time-series momentum.

Goyal and Jegadeesh (2018) compare the performance of the cross-sectional (CS) and time-series (TS) momentum strategies implemented on individual stocks in the USA. They document significantly positive excess returns on both strategies, but the profitability of the TS strategy appears to be larger and stronger. After adjusting for TS excess returns, the alpha of the CS strategy turns to significantly negative, while the alpha of the TS strategy remains strongly positive after adjusting for CS momentum. They note that the key difference between the CS and TS strategies pertains to the threshold of taking a long or short position on financial securities: zero excess return in the case of TS and average excess return in the case of CS. The difference in the threshold results in the difference in proportions of long and short positions in the two strategies. The CS strategies are zero-net investment strategies with equal proportion of long and short positions. In contrast, the TS strategies take on net long position in risky assets that on average earn positive returns.

To make the two strategies comparable, they add to the CS strategy an investment in the market portfolio equal to the dollar value of the difference in long and short sides of the TS strategy. The adjusted-CS strategy performs as well as does the TS strategy and its profit is no longer subsumed by the TS momentum. The result suggests that the superior performance of the TS strategy is due to its net long position in individual stocks. Their results pose a challenge to MOP's (2012) claim that the TS momentum factor may play important roles in explaining asset returns incremental to the CS momentum factor. In short, in contrast to the well-established cross-sectional momentum effect documented by prior studies, the strength of the newly found time-series momentum effect is still a controversial topic in the literature.

Several recent papers study price movements at high frequency and document intraday momentum. For instance, Gao, Han, Li, and Zhou (2018) find that the first half-hour return on the S&P 500 ETF is a strongly positive predictor of its last half-hour return from 1993 to 2013. The predictability is stronger on days with higher return volatility and higher volume, recession days, and days with major macroeconomic news releases. They argue that the intraday market momentum can be explained by a model of infrequent rebalancing via repetitive order activity. The intraday return autocorrelation can also be explained by informed investors' preference to trade in the last half hour to avoid overnight risk or to take advantage of high liquidity. Baltussen, Da, Lammers, and Martens (2021) document strong intraday momentum in returns on more than 60 futures on equities, bonds, commodities, and currencies from 1974 to 2020. Specifically, the return from the previous market close to the last 30 min strongly positively predict the return in the last 30 min before the market closes, while the positive predictability seems to revert in the next few days. They argue that the intraday momentum is driven by the hedging needs of market makers due to short gamma exposures. In particular, to maintain their positions delta-neutral, market makers need to purchase (sell) underlying assets when prices are rising (falling), resulting in price momentum. Lou, Polk, and Skouras (2019) provide a novel decomposition of stock returns into overnight and intraday components to examine the role of investor heterogeneity in a variety of asset pricing

anomalies. They find that the four momentum strategies including the price, industry, earnings, and time-series momentum all earn profits overnight. Moreover, institutional investors appear to on average trade against momentum during the day, suggesting that momentum profits are mainly driven by overnight trading of retail investors.

46.2 The Implementation of Momentum Strategies

This section introduces the detailed procedures to implement the two types of momentum strategies.

46.2.1 Cross-Sectional Momentum Strategies

The cross-sectional momentum strategy, which is the main focus of the literature, trades securities based on their relative historical performance. We use a momentum strategy that is based on the past 6-month performance with a 6-month holding period as an illustration. Specifically, at the end of each month, all securities in each of the samples are ranked in ascending order based on the past 6-month cumulative returns with dividends. The securities in the bottom 10% (or 20% or 30%) are assigned to the loser (denoted as “L”) portfolio, while those in the top 10% (or 20% or 30%) are assigned to the winner (denoted as “W”) portfolio. These portfolios are value-weighted using the market capitalization of the security at the end of the ranking month as the weight. Each of these portfolios is held for 6 months.

To reduce the effect of nonsynchronous trading and the bid-ask bounce, Jegadeesh and Titman (1993) suggest that we measure returns on these portfolios 1 month after the ranking takes place. If a security has any missing returns during the holding period, we replace them with the corresponding value-weighted market returns. If the returns on the security are no longer available, we rebalance the portfolio in the month the security is deleted from our database. Excess returns on a security are calculated as the returns on that security minus the risk-free rate, which we assume is equal to the 1-month government short-term rate, such as the US Treasury bill rate.

To increase the power of the tests, we construct overlapping portfolios. The winner (loser) portfolio is an overlapping portfolio that consists of the “W” (“L”) portfolios in the previous 6 months. The returns on the winner (loser) portfolios are the simple average of the returns on the six “W” (“L”) portfolios. For instance, the January return on the winner portfolio is the simple average of the six January returns on the “W” portfolios that are constructed from at June to at November in the previous year. The momentum portfolio is the zero-cost, winner-minus-loser portfolio.

46.2.2 Time-Series Momentum Strategies

A time-series momentum strategy is implemented in a similar way. The only difference is that the position of a financial security in a month is determined by its own past return in excess of the risk-free rate. MOP (2012) considers a momentum strategy that is based on the past k month performance with a holding period of h months. At the end of each month, they go long (short) a security if it earns a positive (negative) excess return in the previous 12 months (i.e., $k = 12$) and hold the position for 1 month (i.e., $h = 1$). The return in month t on the momentum strategy represents the simple average of returns on all trading positions chosen from the end of month $t - k - 1$ to the end of month $t - 1$. Specifically, for a financial security in a month t , they decide the position at the end of the previous month $t - 1$ based on its excess return from month $t - k - 1$ to month $t - 1$ and compute the return on the portfolio in month t . Then they move backward for a month and calculate month t return on the portfolio formed by the financial security's excess return from month $t - k - 2$ to month $t - 2$. This process is repeated until they move to the end of month $t - h$ and calculate month t return on the portfolio formed by the excess return from month $t - k - h$ to month $t - h$. The return of the momentum trading strategy in month t with k -months lookback periods and h -months holding period is the arithmetic average of month t returns of the $h - 1$ portfolios.

46.3 Explanations for Momentum Profits

Jegadeesh and Titman (2001) discuss three potential explanations for the profitability of momentum strategies and examine the performance of momentum portfolios over longer horizons in order to differentiate between these hypotheses. The three explanations include: (1) stock prices underreact to information, (2) there is a delayed overreaction to information, and (3) the profits are generated from cross-sectional differences in expected returns.

The first two explanations are consistent with some recent behavioral models. For example, the underreaction explanation is consistent with the Barberis, Shleifer, and Vishny (1998) model where a "conservatism bias" can lead investors to underreact or underweight new information. In the case with a pure conservatism bias, once the information is fully incorporated in prices, there is no predictability in stock returns. In this case, the expected post-holding period returns are zero. Grinblatt and Han (2005) argue that the tendency of investors to hold on to losing stocks and sell winning stocks too soon, driven by the prospect theory and mental accounting (PT/MA), can cause underreaction to information. Such underreaction creates spreads between market prices and fundamental values of stocks, with past winners undervalued and past losers overvalued, resulting in the momentum effect. After controlling for a proxy for aggregate unrealized capital gains, the momentum anomaly disappears.

Da, Guren, and Warachka (2014) develop and test the frog-in-the-pan (FIP) hypothesis that attributes return momentum to investors' limited attention to

continuously arriving information in small amounts. In particular, they propose a behavior model under in which investors underreact to signals with magnitudes lower than a threshold. The model predicts that such delay in response to information would induce persistence in returns, with the persistence more pronounced among stocks with more continuous and small information and weaker among stocks with more discrete and dramatic information. They construct a proxy for information discreteness (ID), defined as the product of the sign of a stock's cumulative return over the past 12 months and the difference in the percentage of days during the formation period with negative and positive returns. A high ID for past winners (losers) indicates that the positive (negative) cumulative return in the past mainly comes from a few positive (negative) daily returns with large magnitudes and more discrete information. Consistent with their FIP hypothesis, they find that momentum profits decrease monotonically among stocks with higher ID.

There are a number of behavioral models that are consistent with a delayed overreaction. Barberis, Shleifer, and Vishny (1998) discuss this possibility and describe what they call the "representative heuristic," which suggests that investors may overly extrapolate a firm's past extraordinary earning growths into the future and hence overreact to positive (or negative) information that is preceded by positive (or negative) information. In addition, Daniel, Hirshleifer, and Subrahmanyam (1998) argue that delayed overreaction can arise because of "self-attribution (or cognitive) bias." That is, investors tend to become more overconfident when their stock picks become winners and take more aggressive positions that push up the prices of winners above their fundamental values. Finally, Hong and Stein (1999) propose a model with two groups of investors: informed investors and technical traders, who do not fully take into account the actions of each other. As a result, information is incorporated slowly into stock prices, providing a potential profit opportunity for technical traders. These traders, however, tend to push prices of past winners above their fundamental values. In each of these behavioral models, prices tend to eventually overreact to information and then reverse when prices eventually revert to their fundamentals. All these behavioral models predict the expected post-holding period returns to be negative.

The third explanation is consistent with an efficient market hypothesis where stocks have different expected rates of return because of different risk exposures. In particular, Conrad and Kaul (1998) emphasize that there would be some evidence of momentum even if there were no time-series variation in expected returns since stocks with high (low) expected returns would be expected to have the highest (lowest) returns in adjacent periods. Berk, Green, and Naik (1999) argue that stock return momentum is a result of persistence in expected returns, as firm risk evolves slowly when old projects are replaced with new ones. Johnson (2002) introduces a model with a stochastic cash flow growth rate to provide a rational explanation for momentum profits. In the model, the high curvature in the nonlinear relation between growth rates and returns results in a positive relation between expected growth rate and growth rate risk. Firms with more positive (negative) price moves are more likely to have higher (lower) growth rate shocks and growth rate levels. As a result, momentum sorts tend to sort stocks by positively priced growth rate risk.

Kelly, Moskowitz, and Pruitt (2021) attribute profitability of momentum to its capability to capture time-varying conditional risk exposures. Specifically, they find that momentum is a significant predictor of future market betas and factor loadings on other risk factors in the Fama and French (2015) five-factor model. They use the instrumental principal component analysis (IPCA) to estimate latent risk factors, to which stocks' exposures depend on time-varying firm characteristics. The conditional factor models estimated via IPCA can largely explain the momentum profits, as the alpha of the momentum strategy relative to the dynamic asset-pricing model is negative and statistically insignificant. Risk-based explanations usually suggest that the profits from a momentum strategy should be the same in post-ranking period.

To test these competing hypotheses, we normally examine the post-holding period returns of momentum portfolios beyond the first year after formation, typically up to 5 years. The empirical evidence from the US (Jegadeesh and Titman 2001) and Asian markets (Chui et al. 2002) appears to support the delayed overreaction explanation. That is, the returns on the momentum portfolio eventually reverse to negative 2–5 years after formation. In addition, Fama and French (1996) find that the Fama-French (1993) three factors cannot explain the momentum profits in the USA. Introducing the profitability and investment factors as in Fama and French (2015) hardly provides any help and still leaves the momentum effect largely unexplained (Fama and French 2016).

Another approach to distinguish the behavior- and risk-based explanations for momentum profits is to quantify the relative contributions of expected and unexpected components of returns on the momentum strategies. While a rational framework suggest that momentum profits should be driven by persistence in expected returns, behavior models indicate that information shocks should play dominating roles. Mao and Wei (2014) adopt a return decomposition approach to distinguish the two competing explanations for cross-sectional price momentum in the US market. Specifically, they use the residual income model and analyst earnings forecasts to decompose individual stock returns into three parts: the expected return component, the cash flow (CF) news component, and the discount rate (DR) news component. They then investigate the contribution of each component to return spreads between past winners and losers. They find that momentum profits are largely driven by a persistent and large CF news component partially offset by a reverse in DF news component, while the expected return component plays virtually no role. The findings seem to favor the behavior-based explanation that investors underreact to positive (negative) CF news of past winners (losers) and reject the risk-based explanation for momentum profits. Lochstoer and Tetlock (2020) use the VAR approach to decompose five well-known anomalies including the momentum anomaly into CF and DR shocks. They also find that CF shocks account for the majority of anomaly returns in the US individual stock market.

Aside from investors' irrational behaviors and cross-sectional differences in expected returns, some studies show that market frictions can also help explain momentum profits. For instance, Lesmond, Schill, and Zhou (2004) find that momentum strategies require frequent trading of stocks and that momentum profits mainly come from stocks with very high trading costs. They estimate various trading

costs including taxes, short sale costs, and bid-ask spreads of momentum strategies, and find that the magnitude of the trading costs is comparable with the strategies' profits. They conclude that momentum profits no longer exist after accounting for the execution costs.

46.4 Momentum Profits Across Firms, Countries, and Times

In addition to the average magnitude of momentum profits, whether and how the profitability of momentum strategies varies across firms, countries, and time periods also attracts large interest in the literature. Investigating the cross-sectional and time-series variations can help us better understand the momentum anomaly and distinguish competing theories to explain it.

46.4.1 Momentum Profits and Firm Characteristics

Firm characteristics such as book-to-market ratio, market capitalization, and turnover are shown to have the ability to predict the cross section of expected stock returns in the USA. Several behavior and rational-based models also predict that momentum profits are related to firm characteristics.

The overconfidence model by Daniel, Hirshleifer, and Subrahmanyam (1998) suggest that momentum profits arise because investors are overconfident. Daniel and Titman (1999) argue that overconfidence is likely to influence the perception of investors relatively more, when they analyze fairly vague and subjective information. They use book-to-market ratio as a proxy for information vagueness. Consistent with their hypothesis, they find that momentum profits are negatively related to the firm's book-to-market ratio in the US market. Chui, Titman, and Wei (2002) find similar results for Asian markets. Sagi and Seasholes (2007) also find that momentum profits are enhanced using stocks with high market-to-book (thus low book-to-market) ratios. They argue that the difference in momentum profits of value and growth stocks can be explained by a rational model under which a firm is a portfolio of revenues from existing assets, operating costs, growth options, and shutdown options. Specifically, firms that perform well in the past are better poised to exploit growth options that are risky and would drive up firm risk, resulting in higher expected returns. To the extent that the book-to-market ratio is an inverse proxy for firms' growth options, the return momentum should be stronger for growth firms with a low book-to-market ratio. Their model also predicts that momentum profitability increases with revenue volatility and decreases with operating leverage, which they confirm in the data.

Trading volume or turnover could also proxy for information vagueness. As suggested by asymmetric information models (see, for example, Blume, Easley, and O'Hara 1994), trading volume reflects investors' disagreement on a stock's intrinsic value. The vaguer the information used to value the firm, the more disagreement among the investors, and hence, the greater the trading volume.

Therefore, the momentum effect should be stronger for firms with high trading volume or turnover. Lee and Swaminathan (2000) find that momentum profits are indeed higher for firms with high turnover ratios in the US market. Chui, Titman, and Wei (2002) also find similar results for Asian markets. Lesmond, Schill, and Zhou (2004) argue that the empirical finding that momentum profits are larger and stronger among small firms with high turnover are also in line with their transaction cost-based explanation for the momentum anomaly.

In contrast, Hong and Stein (1999) predict that stocks with slow information diffusion should exhibit stronger momentum. Hong, Lim, and Stein (2000) provide tests that support this prediction. In particular, except for the smallest decile stocks, the profitability of momentum investment strategies declines sharply with firm size. Hong, Lim, and Stein (2000) also look at momentum profits and analyst coverage and find that holding size fixed, momentum strategies work better for stocks with lower analyst coverage. In addition, they find that the effect of analyst coverage is greater for stocks that are past losers than for stocks that are past winners. They conclude that their findings are consistent with the gradual-information-diffusion model of Hong and Stein (1999).

46.4.2 Momentum Profits and Country Characteristics

To the extent that physiological and cognitive biases are correlated with cultural differences, behavior models also suggest cross-country differences in momentum profits. Chui, Titman, and Wei (2010) use international equity data to test the role of investors' physiological biases in differences in momentum profits across countries. Specifically, they use the individualism index constructed by Hofstede (2001), which is based on survey studies in 50 countries and reflects the extent of people's emphasis on their internal attributes, to measure the degree of investors' overconfidence and self-attribution bias. They find that momentum profits differ across countries, to which a large extent can be explained by cross-country differences in the individualism index. In particular, they sort countries by the individualism index and find that the average monthly return on the momentum strategy that longs winners and shorts losers is 0.6% higher in the top 30% countries with high individualism indexes than in the bottom 30% countries with low individualism indexes. In a regression of momentum profits against the individualism index, the regression coefficient on the individualism index is statistically significant positive. The impact of individualism on momentum profits is robust to controlling for other country-specific variables including information uncertainty and measures related to development and integrity of financial markets. Thus, they find evidence to support the implication of the behavior models that the momentum strategy should be more profitable in countries with more individualistic cultures.

46.4.3 Momentum Profits and Market States

Another strand of the literature investigates time-series variations in the profitability of momentum strategies. Many empirical studies document that momentum profits are pro-cyclical. For the sample period from 1929 to 1995 in the US market, Cooper, Gutierrez, and Hameed (2004) find that the average return of the momentum strategy is negative following market downturns, while it is positive following market upturns, with the up-market momentum eventually reverses in the long run. They argue that the behavior model of Daniel, Hirshleifer, and Subrahmanyam (1998) incorporating investors' overconfidence can explain the cross-states differences in momentum profits. Increasing market prices may result in greater aggregate overconfidence, leading to the more pronounced momentum effect. Besides, the pro-cyclicality of momentum profits also seems to be consistent with the implication of Hong and Stein (1999)'s model with two types of agents. In the model, the degree of delayed overreaction increases with decreasing risk aversion. To the extent that risk aversion decreases with increasing wealth, the model predicts higher momentum profits in a bull market.

Stivers and Sun (2010) investigate the intertemporal relation between the cross-sectional dispersion in stock returns (RD) and subsequent value and momentum premiums in the USA. They find that RD is a strongly negative predictor of momentum profits, indicating that the momentum strategy is less profitable in volatile markets. They argue that pro-cyclical momentum profits can be explained by Johnson's (2002) model under which growth rate levels are positively related to growth rate risk, as the expected cash flow growth rate is higher in a bull market. The negative relation between RD and profitability of momentum strategies is also in line with the implication of Sagi and Seasholes's (2007) model accounting for firms' growth options, because firms are closer to exercise their growth options when the market is up.

More recently, Daniel and Moskowitz (2016) document "momentum crashes" in the US equity market, which refer to infrequent and persistent losses of momentum strategies that long past winners and short last losers (WML). The occasional crashes usually occur following economic downturns and contemporaneous with market rebounds. Besides, they find that the momentum crashes mainly come from the short side of the WML strategy, as extreme losers tend to experience substantial return reversal after a financial crisis. They argue that the momentum crashes can be explained by time-varying market betas of momentum strategies and options-like payoffs of loser portfolios during financial downturns. Specifically, they find that loser portfolios mainly consist of highly levered stocks with high default probability. During market recessions, such stocks behave like out-of-the-money call options on their firm values, and tend to have low down-market betas while very high up-market betas. The net effect is that the WML strategy would have very negative market exposures following bear markets and effectively behaves like a short call option on the market. However, the market prices do not adequately reflect the value of such

options, resulting in a high expected return on past losers and a low expected return on the momentum strategy. Consistent with the optionality explanation for the momentum crashes, they find that future momentum returns are negatively correlated with lagged market variance, as options values increase with market variance.

References

- Asness, Clifford S., John M. Liew, and Ross L. Stevens. 1996. Parallels between the cross-sectional predictability of stock returns and country returns. *Journal of Portfolio Management* 23: 79.
- Baltussen, Guido, Zhi Da, Sten Lammers, and Martin Martens. 2021. Hedging demand and market intraday momentum. *Journal of Financial Economics*, forthcoming.
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. A model of investor sentiment. *Journal of Financial Economics* 49: 307–343.
- Berk, Jonathan B., Richard C. Green, and Vasant Naik. 1999. Optimal investment, growth options, and security returns. *Journal of Finance* 54: 1553–1607.
- Blume, Lawrence, David Easley, and Maureen O'Hara. 1994. Market statistics and technical analysis: The role of volume. *Journal of Finance* 49: 153–181.
- Chan, Kalok, Allaudeen Hameed, and Wilson Tong. 2000. Profitability of momentum strategies in the international equity markets. *Journal of Financial and Quantitative Analysis* 35: 153–172.
- Chui, Andy C.W., Sheridan Titman, and K.C. John Wei. 2002. Momentum, legal system, and ownership structure: An analysis of Asian stock markets. Working paper, University of Texas at Austin.
- . 2010. Individualism and momentum around the world. *Journal of Finance* 65: 361–392.
- Conrad, Jennifer, and Gautam Kaul. 1998. An anatomy of trading strategies. *Review of Financial Studies* 11: 489–519.
- Cooper, Michael J., Roberto C. Gutierrez Jr., and Allaudeen Hameed. 2004. Market states and momentum. *Journal of Finance* 59: 1345–1365.
- Da, Zhi, Umit G. Gurun, and Mitch Warachka. 2014. Frog in the pan: Continuous information and momentum. *Review of Financial Studies* 27: 2171–2218.
- Daniel, Kent, and Tobias J. Moskowitz. 2016. Momentum crashes. *Journal of Financial Economics* 122: 221–247.
- Daniel, Kent D., and Sheridan Titman. 1999. Market efficiency in an irrational world. *Financial Analysts Journal*, November/December, 28–40.
- Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam. 1998. Investor psychology and security market under- and overreactions. *Journal of Finance* 53: 1839–1886.
- Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56.
- . 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance* 51: 55–84.
- . 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116: 1–22.
- . 2016. Dissecting anomalies with a five-factor model. *Review of Financial Studies* 29: 69–103.
- Gao, Lei, Yufeng Han, Sophia Zhengzi Li, and Guofu Zhou. 2018. Market intraday momentum. *Journal of Financial Economics* 129: 394–414.
- Georgopoulou, Athina, and Jiaguo Wang. 2017. The trend is your friend: Time-series momentum strategies across equity and commodity markets. *Review of Finance* 21: 1557–1592.
- Goyal, Amit, and Narasimhan Jegadeesh. 2018. Cross-sectional and time-series tests of return predictability: What is the difference? *Review of Financial Studies* 31: 1784–1824.
- Grinblatt, Mark, and Bing Han. 2005. Prospect theory, mental accounting, and momentum. *Journal of Financial Economics* 78: 311–339.
- Grundy, B. D., and Martin, J. S. M. 2001. Understanding the nature of the risks and the source of the rewards to momentum investing. *Review of Financial Studies* 14(1): 29–78.

- Hofstede, Geert. 2001. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Beverly Hills: Sage.
- Hong, Harrison, and Jeremy C. Stein. 1999. A unified theory of underreaction, momentum trading and overreaction in asset markets. *Journal of Finance* 54: 2143–2184.
- Hong, Harrison, Terence Lim, and Jeremy C. Stein. 2000. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance* 55: 265–295.
- Huang, Dashan, Jiangyuan Li, Liyao Wang, and Guofu Zhou. 2020. Time series momentum: Is it there? *Journal of Financial Economics* 135: 774–794.
- Jegadeesh, Narasimhan, and Sheridan Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48: 65–91.
- . 2001. Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance* 56: 699–720.
- Johnson, Timothy C. 2002. Rational momentum effects. *Journal of Finance* 57: 585–608.
- Kelly, Bryan T., Tobias J. Moskowitz, and Seth Pruitt. 2021. Understanding momentum and reversal. *Journal of Financial Economics*, forthcoming.
- Kim, Abby Y., Yuman Tse, and John K. Wald. 2016. Time series momentum and volatility scaling. *Journal of Financial Markets* 30: 103–124.
- Lee, Charles M.C., and Bhaskaran Swaminathan. 2000. Price momentum and trading volume. *Journal of Finance* 55: 2017–2069.
- Lesmond, David A., Michael J. Schill, and Chunsheng Zhou. 2004. The illusory nature of momentum profits. *Journal of Financial Economics* 71: 349–380.
- Lewellen, Jonathan. 2002. Momentum and autocorrelation in stock returns. *Review of Financial Studies* 15: 533–564.
- Lim, Bryan Y., Jiaguo George Wang, and Yaqiong Yao. 2018. Time-series momentum in nearly 100 years of stock returns. *Journal of Banking & Finance* 97: 283–296.
- Liu, Shuming, and K.C. John Wei. 2004. *Do industries explain the profitability of momentum strategies in European markets?* Working paper, Hong Kong University of Science & Technology.
- Lochstoer, Lars A., and Paul C. Tetlock. 2020. What drives anomaly returns? *Journal of Finance* 75: 1417–1455.
- Lou, Dong, Christopher Polk, and Spyros Skouras. 2019. A tug of war: Overnight versus intraday expected returns. *Journal of Financial Economics* 134: 192–213.
- Mao, Qinghao Mike, and K.C. John Wei. 2014. Price and earnings momentum: An explanation using return decomposition. *Journal of Empirical Finance* 28: 332–351.
- Moskowitz, Tobias J., and Mark Grinblatt. 1999. Do industries explain momentum? *Journal of Finance* 54: 1249–1290.
- Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen. 2012. Time series momentum. *Journal of Financial Economics* 104: 228–250.
- Nijman, Theo, Laurens Swinkels, and Marno Verbeek. 2004. Do countries or industries explain momentum in Europe? *Journal of Empirical Finance* 11: 461–481.
- Richards, Anthony J. 1997. Winner-loser reversals in national stock market indices: Can they be explained? *Journal of Finance* 52: 2129–2144.
- Rouwenhorst, K. Geert. 1998. International momentum strategies. *Journal of Finance* 53: 267–284.
- . 1999. Local return factors and turnover in emerging stock markets. *Journal of Finance* 55: 1439–1464.
- Sagi, Jacob S., and Mark S. Seasholes. 2007. Firm-specific attributes and the cross-section of momentum. *Journal of Financial Economics* 84: 389–434.
- Stivers, Chris, and Licheng Sun. 2010. Cross-sectional return dispersion and time variation in value and momentum premiums. *Journal of Financial and Quantitative Analysis* 45: 987–1014.



Equilibrium Credit Rationing and Monetary Non-neutrality in a Small Open Economy 47

Ying Wu

Abstract

This chapter modifies the well-known Mundell-Fleming model by adding equilibrium credit rationing as well as imperfect asset substitutability between bonds and loans. When the representative bank's backward-bending loan supply curve peaks at its profit-maximizing loan rate, credit rationing can be an equilibrium phenomenon, which makes credit-dependent capital investment solely dependent upon the availability of customer-market credit. With the rationing, an expansion in money and credit shifts the IS curve as well as the LM curve even in a small open economy under a regime of fixed exchange rates, and the magnitude of offset coefficient between domestic and foreign asset components of high-powered money is less than one. In contrast, if there is no credit rationing, imperfect asset substitutability between bonds and loans per se cannot generate the real effect of money in the same model.

Keywords

Credit rationing · Monetary policy · Capital flow · Mundell-Fleming model · Monetary neutrality · Open market operation · IS-LM curves · Offset coefficient · Monetary base · Small open economy

JEL Classification

E51 F41

Y. Wu (✉)

Department of Economics and Finance, Franklin P. Perdue School of Business, Salisbury University, Salisbury, MD, USA
e-mail: yxwu@salisbury.edu

Abbreviations

IS	Investment-saving (curve)
LM	Liquidity-money (curve)
BP	Balance of payment (curve)

47.1 Introduction

Is money neutral in a small open economy with international capital mobility and a fixed exchange rate regime? Can monetary policy affect real output in the circumstances? The answer to these questions is widely construed to be negative because the money supply has lost its role of a nominal anchor in this case.¹ In the orthodox money view, it is the interest rate that serves as the channel through which monetary policy affects the real sector of an economy; however, because the interest rate channel of monetary policy is highly correlated with exchange rates, and because the monetary authority commits to the maintenance of the fixed exchange rate, the consequent foreign exchange intervention by the monetary authority using official reserves necessarily washes out any real effect of the monetary policy that it has previously initiated. The same approach used in most of the existing literature on small open economies, such as the traditional IS/LM analysis, holds a lopsided view of bank liabilities and bank loans: other than influencing interest rates via manipulating deposits (a money asset and bank liability), banks have no active leverage to play with; the role of bank loans escapes unnoticed since bank loans are grouped together with other nonmonetary assets such as bonds.

In contrast to the money view, the credit view of monetary transmission mechanism rejects the notion that all nonmonetary assets are perfect substitutes. According to the credit view, due to information asymmetries between borrowers and lenders in financial markets, banks can play a particular role in reducing information costs. It is financial intermediation that can help firm with risk-sharing, liquidity, and information services; as a result, a large number of firms have in fact become bank dependent. Furthermore, although a rise in the loan rate increases, *ceteris paribus*, the bank's expected return by increasing interest payment when the borrower does not default, it lowers the bank's expected return by exacerbating adverse selection and moral hazard problems and thus raising the probability of default. Hence, the bank's loan supply curve can be backward-bending and credit rationing may occur as an equilibrium phenomenon.² Credit rationing *per se* makes monetary credit availability rather than interest rates to be the conduit for the real

¹This line of analysis can be traced back to Fleming (1962) and Mundell (1963).

²In their influential paper, Stiglitz and Weiss (1981) provides information-based analysis of equilibrium credit rationing. Blinder and Stiglitz (1983) further argues that monetary policy works through bank credit for there are no close substitutes for it at least as far as most medium and small firms with relatively high risk are concerned. For a comprehensive review of the credit-view literature, see Kashyap and Stein (1993).

effect of money, therefore providing a major theoretical underpinning for the effectiveness of monetary policy under fixed exchange rates.

This chapter begins with a study of the loan market setting with asymmetric information as a micro-foundation for consumption and investment, and further develops a macro model of a small open economy under a fixed exchange rate regime with perfect capital mobility in the bond market and imperfect asset substitutability between bonds and loans. As far as the credit view is concerned, this chapter in spirit is close to Bernanke and Blinder (1988); both address the credit channel of monetary policy in a variant of the IS/LM model. They differ in several regards, however. Unlike Bernanke and Blinder, the model in this chapter incorporates the possibility of equilibrium credit rationing while maintaining the assumption of imperfect substitutability of bank loans and bonds. With imperfect substitutability between bonds and bank loans, this chapter nests both credit-rationed and credit-unrationed equilibrium regimes. Additionally, by placing the credit channel of monetary policy in the setting of a small open economy, this chapter allows the possibility to explore the relevance of the monetary-policy-ineffectiveness proposition in the existing mainstream small-open-economy literature.

Partly based on Wu (1999) by drawing on its microeconomic foundation setting, this study has made important and substantial revisions to its macroeconomic analysis. With the credit availability channel, this study shows that money in the fixed exchange rate model is not completely endogenous by appealing to the asymmetry between customer-market credit and auction-market credit under equilibrium credit rationing.³ Incorporating bank credit into the fixed exchange rate model leads to two fundamental changes. First, it extends the scope for monetary policy to affect economy from the standard interest rate channel to the one including the bank lending channel and balance-sheet channel as well; the latter two conduits can be independent of changes in interest rates. Second, and more importantly, monetary policy will no longer be deemed impotent since it can directly “shift” the goods market as well as money market equilibrium schedules in such a way that the targeted real effect could be achieved while the fixed exchange rate is sustained.

In what follows, the next section presents the analytical structure of bank behavior and credit market; the following two sections explore how credit market conditions determine macroeconomic equilibrium in an open-economy IS/LM framework and demonstrate the real impacts of monetary shocks through its credit channel, respectively. The final section concludes the study.

³It is shown indeed in my earlier paper (Wu 1999) that monetary policy can have real effects in both credit-rationing regime and the market-clearing regime if the central bank's foreign exchange reserves are independent of open market operations. While considering the endogeneity of the monetary base, Ramírez (2001) argues that “monetary policy is still ineffective in influencing output under a fixed exchange rate, even with an operative credit channel.” Nevertheless, in response to Ramírez (2001), this chapter shows particularly how credit-rationing channel can save monetary policy from being charged with impotence even in a small open economy with complete capital mobility and fixed exchange rates.

47.2 Bank Behavior and Credit Market

It is well known that due to the credit risk associated with adverse selection and moral hazard problems a banking firm has an inverse U-shaped supply curve of loans with a backward-bending portion. This section essentially modifies the pedagogical model in Christopher and Lewarne (1994) by extending the spectrum of bank investment into the portfolio selection between bonds and loans.

The representative banking firm is assumed to hold exactly the required amount of reserves and allocate all of its excess reserves between the two bank assets: bonds and loans. Thus, it chooses loans, l , subject to its balance-sheet identity, to maximize its profits from lending

$$\Pi = \theta(\rho)l\rho + b_b r - dr - \frac{\gamma}{2}l^2 \quad (1)$$

$$s.t. \quad b_b + l = (1 - k)d$$

where ρ is the loan rate, $\theta(\rho)$ is the probability of loan repayment, γ is the cost parameter of servicing loans, b_b denotes bonds held by the banking firm, r is the interest rate on bond, d represents total deposits, and k is the required reserve ratio for deposits. Here, the low-risk or risk-free interest rate on bond holding is assumed to be the same as the interest cost of taking in deposits. Thus, deposits and bonds are perfectly substitutable assets to depositors so that they pay the same per-dollar expected return. The key characteristic of the bank profit is that the repayment probability depends on the loan rate. Following the existing literature on equilibrium credit rationing, an increase in the loan rate makes it more likely for borrowers to default; hence, the repayment probability is a decreasing function of the loan rate.⁴ In addition, the representative bank takes the flow of deposits as given when making its portfolio decisions. Substituting the balance-sheet identity into the bank's objective function and maximizing it with respect to l yields the banking firm's loan supply curve

$$l^s = \frac{\theta(\rho)\rho - r}{\gamma} \quad (2)$$

Several implications of the loan supply curve can be derived. First, the loan supply curve is backward bending. The co-movement of the loan rate and loan volume hinges on the elasticity of the odds of repayment with respect to the loan rate. Only when the repayment probability is inelastic can exist the positive

⁴As shown by Stiglitz and Weiss (1981) and others, the loan rate an intermediary charges may have both an adverse selection effect and a moral hazard effect on the risk of a pool of loans. Raising the loan rate shifts the mix of borrowers toward riskier firms and their projects financed by the loans, thus reducing the lender's expected return. As a result, the intermediary may maximize its expected profits by setting an interest rate at a level that results in an excess demand for bank credit.

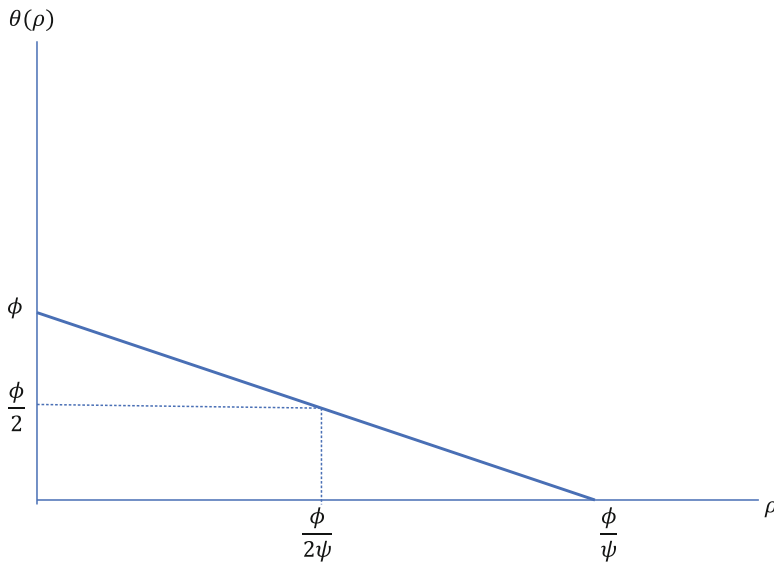


Fig. 1 Loan-repayment probability function

relationship between the loan rate and loan volume. To be specific, consider a linear repayment probability $\theta(\rho) = \phi - \psi\rho$, where ϕ is the autonomous repayment probability determined by noninterest factors such as the liquidity of balance sheet positions, and ψ measures the sensitivity of the repayment probability to the loan rate ($0 < \psi < \phi \leq 1$). Fig. 1 depicts the loan-repayment probability function. Facing the linear loan repayment probability function, the loan volume supplied increases with the loan rate until the loan rate achieves $\phi/2\psi$; after that, a higher loan rate actually reduces the loan volume. In Fig. 1, the loan rate at which the loan supply curve begins to bend backward points to the repayment probability halfway to its maximum in the possible range.

Substituting $\theta(\rho) = \phi - \psi\rho$ into (2) and differentiating (2) with respect to r , ϕ , ψ , and γ produces the responses of loan supply to the parameters of servicing loans. In particular, an increase in the bond interest rate, r , ceteris paribus, makes bond-holding more attractive; accordingly, banks will reduce loans and hold more bonds. Another interpretation for the decrease of bank loans is based on the equivalence between the bond interest rate and the deposit rate: the higher the interest expenses of raising loanable funds by issuing deposits, the higher the economic cost of making loans. Next, banks tend to issue more loans when the autonomous repayment probability, ϕ , is higher due to borrowers' increased net worth, for example. In addition, the larger the sensitivity of the repayment probability to the loan interest rate, ψ , the more deteriorating the problems of adverse selection and moral hazard, thus it is more likely for credit rationing to occur. Finally, an increase in the cost of servicing loans, γ , tends also to reduce loans as long as the expected return per dollar of loans exceeds the corresponding real opportunity cost.

Applying the envelope theorem to the representative bank's profit function in (1) while incorporating (2) and $\theta(\rho) = \phi - \psi\rho$ generates the following marginal bank profit with respect to the loan rate:

$$\frac{d\Pi(\rho)}{d\rho} = \frac{1}{\gamma} [2\psi^2\rho^3 - 3\psi\phi\rho^2 + (2\psi r + \phi^2)\rho - \phi r] \quad (3)$$

The bracket term on the RHS of (3) is a cubic expression but two of the three roots are degenerated solutions at which loans are zero respectively; thus, the only feasible root to (3) is $\rho^* = \phi/2\psi$, at which the bank's expected profits are maximized. Recall that the bank's loan supply curve peaks exactly at the same loan rate as the profit-maximizing loan rate here. Therefore, the result suggests the existence of equilibrium credit rationing. Furthermore, the result for profit-maximizing loans also imply that the loan interest rate exceeds the bond interest rate such that $\rho > \sqrt{r/\psi} > r$, which captures the existence of risk premium of bank lending and therefore signifies the imperfect substitutability between loans and bonds.

Moving from the representative bank to the aggregate banking system, the aggregated bank balance sheet identity shows $B_b + L + R = D$, where B_b is the bonds held by banks, D denotes deposits, and L is the volume of loans. For simplicity, currency is abstracted from the model. The required reserve of the banking system, R , constitutes the monetary authority's liabilities, or high-powered money, H , which are generated by its acquisition of bonds (B_a) and foreign exchange (F). The high-powered money in this framework is composed of exclusively required reserves; the money supply can be expressed by H/k .

Suppose that there are n banks, with the representative bank's supply of loans specified in Eq. (2), aggregating generates the total supply of loans. A structural view of the aggregated balance sheet of banks suggests that if banks allocate a fraction of their excess reserves into loans and the rest into bonds, the aggregate supply of loans is given by $\varepsilon \cdot (1 - k) \cdot (H/k)$, where ε represents the ratio of loans to excess reserves. Accordingly, the share of loans in excess reserves must characterize the banks' loan-making behavior, and it is thus actually a function of the same set of variables that determine aggregate supply of loans.

$$L^S = \varepsilon(\rho, r, \phi, \psi, \gamma, n) \frac{1-k}{k} H \quad (4)$$

? - + - - +

where the symbols underneath each of the arguments in $\varepsilon(\bullet)$ denote the signs of the partial derivatives associated with them. For simplicity, it is assumed that bank credit is the only debt instrument for firms to finance their investment; investment demand and the demand for bank loans are taken to be equal.⁵ Thus, aggregate demand for loans is negatively related to the loan interest rate, and its standard linear form is.

⁵ More generally, the demand side of the loan market is influenced by the interest rates on the two credit instruments, loans and bonds, as well as aggregate income, see Bermanke and Blinder (1988).

$$L^D = \alpha - \beta\rho \quad (5)$$

Indeed, as demonstrated by the existing literature on markets in disequilibrium, the loan market may or may not be at the market-clearing equilibrium.⁶ Nevertheless, unlike the school camp of disequilibrium economics, the loan quantity traded in the market is not uniformly characterized by the minimum of demand and supply sides. Loan rationing can arise in an unrestricted market setting flawed only by plausible information asymmetries; the loan rate can always freely adjust to a level consistent with market forces driven by the profit-maximization incentives. Therefore, credit rationing could exist at the profit-maximizing loan rate, $\rho^* = \phi/2\psi$, and sustain as an equilibrium phenomenon. The excess demand fails to drive the loan rate upward because the associated credit risk would reduce banks' profits; however, if at the same loan rate there is an excess supply, the loan interest rate will adjust downward to clear the loan market since holding excess reserves does not add to profits at all.

Consider the demand for and supply of loans specified in (4) and (5), respectively, the equilibrium interest rate in the loan market is given by

$$\rho = \begin{cases} \frac{\phi}{2\psi} & , \text{ if } L^D \geq L^S \text{ at } \frac{\phi}{2\psi} ; \\ \min(\rho_1, \rho_2 | L^D = L^S) & , \text{ if } L^D < L^S \text{ at } \frac{\phi}{2\psi} . \end{cases} \quad (6)$$

where ρ_1 and ρ_2 are the two roots of the quadratic equation given by $L^D = L^S$. Recall that $\rho^* = \phi/2\psi$ is the loan rate that corresponds to the maximum quantity of loans. If an excess supply exists at ρ^* , L^D must cross L^S once at a loan rate below ρ^* and once at a loan rate above ρ^* . Since ρ^* is the profit-maximizing loan rate, the bank has no incentive to raise the loan rate to any level above ρ^* ; credit is then rationed at the equilibrium. On the other hand, the profit-maximizing loan rate is not attainable if there is excess supply at ρ^* since the bank cannot force the firms to borrow in excess of the amount that maximizes their profits. It follows that if a bank cannot maximize its profit at ρ^* due to deficient demand, the best attainable outcome for the bank is to allow a downward adjustment in the loan rate until the loan market clears. Therefore, the loan quantity traded is at the market-clearing equilibrium level if the market interest rate of loans is below the banks' desired level, ρ^* ; otherwise, it would be supply determined at the profit-maximizing loan rate.

47.3 Macroeconomic Equilibrium

Assuming that investment is solely dependent on the availability of bank credit, and investment demand is equivalent to the demand for loans. Based on the analytical results in the preceding section, there is an implicit positive relationship between the

⁶For studies on disequilibrium markets under price rigidity, see Barro and Grossman (1971), and Muellbauer and Portes (1978), among others.

interest rates on loans and bonds, which can be explicitly expressed as $\rho = \lambda(r)$. If credit demand is not rationed in the loan market, we have $I(\rho) \equiv L^D[\lambda(r)]$, with $I' = L^D \lambda' < 0$; however, with credit rationing, investment demand is totally determined by the aggregate supply of loans.

47.3.1 Case for Credit Rationing

With credit rationing, the quantity of loans effectively traded is given by L^S specified in (4). In this case, the monetary authority can help loosen credit rationing through open market purchases: the nonbank public selling bonds to the monetary authority deposits the proceeds into banks, and the loan supply increases with the deposits. The rationing situation improves and the resulting increase in output increases money demand and thus imposes upward pressure on the interest rate and the exchange value of the domestic currency, which in turn relieves the money market of the adjustment burden resulting from the monetary authority's commitment to the fixed exchange rate under the circumstances of open market purchases. Therefore, following the monetary authority's open market purchases, although there are market forces to purchase foreign bonds and thus leads the monetary authority to sell foreign reserves, the authority's operation on foreign reserves does not fully sterilize its open market operation on domestic bonds so that its net effects are to expand loans and increase output.

Credit rationing enhances originally existed imperfect asset substitutability between bonds and loans; lending to domestic capital investment under rationing is expected to be more preferable to holding foreign bonds. Without losing generality, a thought experiment could be to assume credit movement to be "segmented" in such a way that the goods market takes only credit expansion from open market purchases on domestic bonds whereas it is asset portfolio adjustment rather than disinvestments in real capital goods that responds to any credit contraction from open market sales on foreign bonds.⁷ Hence, grouping (4) with the equilibrium conditions of the goods market and the money market yields the following simple macroeconomic model:

$$Y = C(Y) + \varepsilon \frac{1-k}{k} B_a + X(Y, eP^f) \quad (7)$$

$$B_a + F = k l(Y, r^f) \quad (8)$$

Note that the domestic price level is normalized at unity since price rigidity applies to the short-run macroeconomic model. Besides the derivatives property of L^S stipulated in (4), the other relevant derivatives in the above model satisfy the following conditions: $C > 0$, $X_Y < 0$, $l_Y > 0$, and $l_r < 0$. Equation (7) is the private-

⁷ Although the assumption of perfect capital mobility rules out the possibility of "home bias," which would otherwise explain the asymmetry between holding of domestic assets and holding of foreign assets, it is credit rationing that holds the key for the real effect of monetary policy here.

sector-only IS equation with the presence of the loan market in an open economy, where $C(\cdot)$ is consumption function, $\varepsilon[(1-k)/k]B_a$ is the supply of loans available to investment from the monetary authority's open market purchases, $X(\cdot)$ is net export function, e is the domestic currency price of foreign exchange, and P^f is the foreign price level. Equation (8) represents the "monetary" version of the open-economy LM equation (or the balance-of-payments equation). r^f is the foreign interest rate on bonds, which equals to r by perfect capital mobility in the bond market, and $l(\cdot)$ is the demand for money, increasing in income and decreasing in the interest rate.

There are three endogenous variables when credit rationing exists in the loan market: income Y , loan quantity L , and the international reserve component in the monetary base F . They are determined simultaneously by three equations: (4), (7), and (8). Although money is partly endogenous due to perfect capital mobility in the bond market and the monetary authority's commitment to maintain fixed exchange rates, the endogeneity of money is not complete because of the credit channel of monetary transmission mechanism and thus money is not completely neutral. Changes in the money supply serve to shift not only the LM curve but also the IS curve so that the responsive change in money does not totally wash out the real effect generated by the monetary change associated with open market purchases. Therefore, the credit channel rescues monetary policy from the charge of impotency.

47.3.2 Case for Non-rationing of Credit

In the regime in which loans are not rationed, both the loan quantity and the loan interest rate are endogenous variables in addition to income and the international reserves of the central bank. The general equilibrium system consists of the loan supply Eq. (4), the monetary version of the balance-of-payments Eq. (8), and two other basic equations below:

$$Y = C(Y) + (\alpha - \beta\rho) + X(Y, eP^f) \quad (9)$$

$$\alpha - \beta\rho = \varepsilon(\rho, r, \phi, \psi, \gamma, n) \frac{1-k}{k} (B_a + F) \quad (10)$$

Equation (9) is the standard IS equation; unlike the credit-rationing counterpart (7), the interest rates play a role in the determination of income. In (10), its LHS is the demand for loans, and the RHS the supply of loans. Equations (4), (8), (9), and (10) implicitly determine the equilibrium values of Y , L , ρ , and F .

47.4 Comparative Static Analysis

The present section examines the responses of the equilibrium income, loan quantity, the loan interest rate, and the international reserve component of the monetary base to a monetary shock initiated through, say, open market operation conducted by the monetary authority. These impacts vary with whether credit is rationed.

Consider the credit-rationing model first. Differentiating Eqs. (4), (7), and (8) with respect to L^S , Y , F , and B_a produces the following results:

$$\frac{dL^S}{dB_a} = \frac{kl_Y(\varepsilon \frac{1-k}{k})^2}{1 - C' - X'} > 0 \quad (11)$$

$$\frac{dY}{dB_a} = \frac{\varepsilon \frac{1-k}{k}}{1 - C' - X'} > 0 \quad (12)$$

$$\frac{dF}{dB_a} = -1 + kl_Y \frac{\varepsilon \frac{1-k}{k}}{1 - C' - X'} < 0 \quad (13)$$

Under the fixed exchange rate system, changes in the official international reserves mirror the status of the balance of payments. Starting from the open market purchase on the part of the monetary authority, the money supply (bank deposits) increases multiply through the money multiplier $1/k$ as high-powered money B_a increases. Banks usually tend to make more loans due to an expanded volume of deposits, and the increased loans relax the credit constraint facing the economy so that income rises, as measured in (11) and (12). Finally, transaction demand for money also increases as a result of increased income but the generated money demand via the credit channel this way is less than the initial increase in money supply, that is, the money creation made by the central bank outpaces the growth of money demand. The resulting excess supply of money must be spent on the purchase of foreign goods or financial assets. As the domestic residents exchange their domestic money for foreign money, the central bank loses international reserves and the money account of the balance of payment moves into a deficit ($dF < 0$). The consequent money contraction will sustain until the disequilibrium in the money market disappears and the balance of payments is back to equilibrium at the foreign interest rate, r^f .

Examining (13) suggests that the credit channel per se plays a role to preserve the legacy of monetary policy. The increase in money demand generated from the credit-driven income expansion lessens the adjustment burden fallen upon the official international reserves so that the absolute value of the offset coefficient is less than one and monetary control is not completely lost.⁸ The economics reasoning and the pertinent empirics suggest that the second term on the RHS of (13) is a positive fraction, and in that term a large credit multiplier, $\varepsilon \cdot (1 - k)/k$, serves to reduce the magnitude of the offset coefficient given by (13). Hence, the stronger the credit channel is, the more legacy of monetary policy can be reserved. However, without considering the credit channel, the endogenous change in foreign reserves would

⁸Some studies have provided the evidence for a certain degree of monetary autonomy under fixed exchange rates. For example, Cumby and Obstfeld (1983) and Rennhack and Mondino (1988) find that structural estimates of offset coefficients are less than one. Also, using Granger causality tests for a number of countries, Montiel (1989) and Dowla and Chowdhury (1991), respectively, produce the finding that some domestic financial aggregates like money and credit Granger-cause domestic real output.

completely offset the initial change in the credit brought about by the central bank's open market operations, and then the traditional result of monetary neutrality would follow.

Figure 2 summarizes the analytical results in a four-quadrant diagram. Quadrant I depicts the IS-LM-BP curves in the traditional general equilibrium framework, with the initial credit-rationing equilibrium as shown in Quadrant III, and Quadrant II depicts the linearized implicit function $\rho = \lambda(r)$. Consider an expansionary monetary policy initiated by the open market purchase. The LM curve shifts rightward initially to the position of the dashed line, causing the IS curve to shift in the same direction through the credit channel. This is reflected in Quadrant III by the downward shift of the aggregate loan supply curve, with the loan interest rate remaining at the profit-maximizing equilibrium level. Due to the tight credit market, the resulting increase in loans directly translates into the corresponding increase in income at the full scale, as depicted in Quadrant IV.⁹ The increased money demand mitigates the excess-supply pressure on the money account of the balance of payments, though the equilibrium in the money market and foreign exchange market still entails a reduction of official international reserves held by the central bank. As a result, although the LM curve shifts backward away from its initial post-shock position as given by the dashed line, it does not shift all the way back to LM_0 ; instead, LM^{CR}_1 meets IS^{CR}_1 at Y^{CR}_1 under the circumstances.

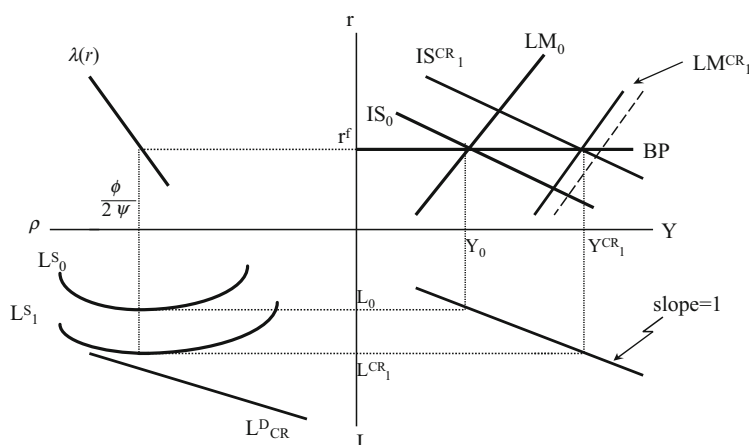


Fig. 2 Monetary non-neutrality under credit rationing in a small open economy with a fixed exchange rate

⁹For simplicity, the income-multiplier effect is ignored here.

Now let us turn to the situation in which there is no credit rationing. Differentiating (4), (8), (9), and (10) with respect to L , Y , ρ , F , and B_a generates the following comparative static results:

$$\frac{dL}{dB_a} = 0 \quad (14)$$

$$\frac{dY}{dB_a} = 0 \quad (15)$$

$$\frac{d\rho}{dB_a} = 0 \quad (16)$$

$$\frac{dF}{dB_a} = -1 \quad (17)$$

As shown by the comparative static results in Eqs. (14) through (17), there is a sharp contrast between the cases of credit rationing and non-rationing of credit in terms of monetary neutrality and the effectiveness of monetary policy. In the absence of credit rationing, the credit channel can only operate through its impact on the loan interest rate. Nevertheless, the loan interest rate is directly related to the world interest rate through $\rho = \lambda(r)$, and the world interest rate cannot be influenced by a small open economy's monetary authority. Following an open market purchase, once the domestic interest rate tends to decline from the level of the world interest rate and thus the exchange rate may deviate from its par accordingly, the monetary authority is obligated to contract money and credit by selling its foreign reserves until the asset prices restore their initial equilibrium levels, therefore rendering the intended monetary expansion ineffective.

47.5 Conclusion

This chapter has demonstrated that equilibrium credit rationing plays a role in preserving the legacy of monetary policy even under a fixed exchange rate regime with perfect international capital mobility. Under equilibrium credit rationing, credit-dependent investment transmits a monetary shock into changes in real income and thus transaction demand for money, therefore sharing the adjustment burden of maintaining the fixed exchange rate, which would otherwise completely fall upon the official international reserves. The magnitude of offset coefficient becomes less than one since any expansion of domestic credit and its real effect is only partially offset by the associated monetary contractions happening through international financial portfolio investment. The degree of retained monetary autonomy depends on the magnitude of the credit multiplier under rationing.

When there exists equilibrium credit rationing, monetary contractions resulting from the monetary authority's endogenous open market sale of foreign assets will take a conduit of portfolio disinvestments rather than real capital disinvestments.

Therefore, the asymmetry between domestic real capital investment (customer-market credit) and financial portfolio investment (auction-market credit) in responding to impulses of open market operations holds the key for monetary non-neutrality.

In contrast, incorporating credit market without credit rationing into analysis fails to rescuer monetary policy from its neutrality in a small open economy committed to fixed exchange rates. That is, the assumption of imperfect substitutability per se between auction-market credit (bonds) and customer-market credit (bank loans) is insufficient for monetary autonomy in the Mundell-Fleming model, though it is adequate for the modified IS-LM type model (Bernanke and Blinder 1988, for example) in which the features of small open economy and fixed exchange rates do not appear.

References

- Barro, R.J., and H.I. Grossman. 1971. A general disequilibrium model of income and employment. *American Economic Review* 61: 82–93.
- Bernanke, B., and A. S. Blinder. 1988. Credit, money, and aggregate demand. *American Economic Review* 78 (Paper and Proceedings): 435–439.
- Blinder, A. S., and J. Stiglitz. 1983. Money, credit constraints, and economic activity. *American Economic Review* (Papers and Proceedings) 73: 297–302.
- Christopher, J.W., and S. Lewarne. 1994. An expository model of credit rationing. *Journal of Macroeconomics* 16 (3): 539–545.
- Cumby, R.E., and M. Obstfeld. 1983. Capital mobility and the scope for sterilization: Mexico in the 1970s. In *Financial policies and the world capital market*, ed. P.A. Armella, R. Dornbusch, and M. Obstfeld, 245–276. Chicago: University of Chicago Press.
- Dowla, A., and A. Chowdhury. 1991. Money, credit, and real output in the developing economies. Unpublished. Department of Economics, Marquette University.
- Fleming, J.M. 1962. Domestic financial policies under fixed and floating exchange rates. *International Monetary Fund Staff Papers* 9 (November): 369–379.
- Kashyap, A., and J. Stein. 1993. Monetary policy and bank lending. *NBER Working Paper* no. 4317.
- Montiel, P.J. 1989. Empirical analysis of high-inflation episodes in Argentina, Brazil and Israel. *IMF Staff Papers* 36 (September): 527–549.
- Muellbauer, J., and R. Portes. 1978. Macroeconomic models with quantity rationing. *Economic Journal* 88: 788–821.
- Mundell, R.A. 1963. Capital mobility and stabilization under fixed and flexible exchange rates. *Canadian Journal of Economics and Political Science* 29 (November): 475–485.
- Ramírez, C.D. 2001. Even more on monetary policy in a small open economy. *International Review of Economics and Finance* 10 (3): 399–405.
- Rennhack, R., and G. Mondino. 1988. Capital Mobility and Monetary Policy in Colombia. *Working Paper* no. 88/77, International Monetary Fund (August).
- Stiglitz, J., and A. Weiss. 1981. Credit rationing in markets with imperfect information. *American Economic Review* 71: 393–410.
- Wu, Y. 1999. More on monetary policy in a small open economy: A credit view. *International Review of Economics and Finance* 8 (2): 223–235.



Policy Coordination Between Wages and Exchange Rates in Singapore

48

Ying Wu

Abstract

Singapore's unique experience in macroeconomic management involves the government's engagement in a tripartite collective bargaining and its influence on the macroeconomic policy game in wages and exchange rates in response to inflation and output volatility. The period from the mid-1980s to mid-1990s features the policy game with a Nash equilibrium in the level of wages and exchange rates and a non-Nash equilibrium in wage growth and exchange rate appreciations. Based on the empirical evidence in the period, the models used in this study suggest that wage and exchange-rate police are a pair of complements both at their levels (Nash equilibrium) and at their percentage changes (non-Nash equilibrium).

Keywords

Effective exchange rates · Collective bargaining · Nash equilibrium · Unit labor cost National Wages Council · Monetary Authority of Singapore

JEL Classification

E61 · F41 · E51

Abbreviations

NWC	National Wages Council
MAS	Monetary Authority of Singapore
CPF	Central Provident Fund
VE	Vector Error Correction

Y. Wu (✉)

Department of Economics and Finance, Franklin P. Perdue School of Business, Salisbury University, Salisbury, MD, USA

e-mail: yxwu@salisbury.edu

LULC	Logarithm of Unit Labor Costs
LNEER	Logarithm of Nominal Effective Exchange Rate;
LIMP	Logarithm of Import Price

48.1 Introduction

Adverse supply shocks often pose a dilemma for the Keynesian approach to aggregate demand management: Implementing expansionary monetary and fiscal policies tends to exacerbate inflation whereas the laissez-faire policy stance is conducive to acute and prolonged unemployment before the economy restores its natural rate level of output. As an alternative means to avoid the predicament and cope with demand shocks as well as supply shocks, appropriate labor market policies including wage policy are recently gaining their importance in macroeconomic management.¹ Nevertheless, wages tend to be sticky downward, and it becomes difficult to attempt to reduce them due to the existence of strong labor unions or the laws prohibiting such measures. The idea of instituting an agreement by unions and corporations to link wage growth with productivity growth, though attractive, often faces great political and economic challenges when it is put in practice.² Accordingly, in general there is a dearth of research on the effectiveness of wage policy in an environment where other aggregate demand policies exist.

Singapore is an ideal case for the study of the effectiveness and dynamics of wage and exchange-rate policies, not only because it has actively deployed wage policy in combination with exchange-rate policy for more than two decades but also because it has maintained a remarkable record of sustained economic growth with low inflation in a small open economy.³ As a highly opened small economy, Singapore faces the challenges of “imported” foreign inflation as well as the wage-push inflation that results from rapid economic growth and labor shortage. The exchange rate and wage movements naturally become the two inter-related key factors in maintaining macroeconomic stability. Specifically, the wage policy manipulated by a tripartite collective bargaining institution known as the National Wages Council (NWC) has

¹In the euro area, particularly, the familiar policy instruments such as the exchange rate and money supply have ceased to be available at the national level while fiscal policy is also often constrained by the straitjacket that the budget deficit cannot exceed 3% of GDP, which renders more room for national wage policies (Calmfors 1998; Wu 1999; Lawler 2000; Karadeloglou et al. 2000; Abraham et al. 2000).

²The Council of Economic Advisers to the President in the United States explicitly implemented income policies by imposing the general guidepost for wages from 1962 to 1965 (see Perry 1967, and Schultz and Aliber 1966). The guidepost implicitly remained in practice from time to time in the 1970s as well. In the United Kingdom, the 1980s and 1990s saw a resurgence of interest in income policies due to rising unemployment. For an argument for wage policy, see Hahn (1983, p.106).

³The average annual GDP growth in Singapore over the last decade was greater than 7.5%, with an inflation rate of about 2% per year.

actually acted as an important complement to the country's exchange-rate policy controlled by the Monetary Authority of Singapore (MAS) (Otani and Sassanpour 1988; Wu 1999).

The NWC is made up of representatives from the government, labor unions, and employer federations. Its main function is to select the wage policy that is not only agreeable by all three parties but also compatible with macroeconomic targets. Although the NWC's wage recommendations only sketch a guideline for negotiations between employers and employees, both public and private sectors usually accept and implement them rather smoothly. The resulting collective bargaining agreements often extend to nonunion workers as well. Labor unions in Singapore actively promote sound economic policies to their members and support restraints when needed. In this way, the wage council helps to reduce the frictions that information asymmetry and costly bargaining often cause in supply-side adjustments.⁴ In coordination with the NWC's endeavor in achieving orderly wage settlements, the Monetary Authority of Singapore (MAS), as the other key player in Singapore's macroeconomic management, chooses the optimal exchange rate variation to cope with the dual inflationary pressures (i.e., the imported inflation, and the inflation pushed by labor shortages) and to maintain the economy's competitiveness.

With its focus on the role of collective bargaining in macroeconomic management in Singapore, this study attempts to model the policy game of wage and exchange rate policies between the NWC and the MAS. The study starts with an analysis of the behavior of wage and exchange rate levels in the policy game and its empirics. It then further derives the MAS' exchange-rate response function and the NWC's wage response function, in terms of the two policies' percentage changes, and analyzes two interplay patterns of the two response functions: the Nash game and the non-Nash game. For the non-Nash game, the study calibrates the analytical outcome in each of the three potential scenarios of the economy: inflation, recession, and the "Goldilocks" scenario (neither inflationary nor recessionary) and compares the simulation results with the actual quarterly growth paths under the two policy rules for the period from 1987:1 to 1996:4.

⁴Singapore's system of national wage council has distinguished itself from the centralized collective bargaining in European countries in three aspects. First, unlike the intermittent European government involvement in wage negotiations, the Singapore government has continuously committed to its participation in the yearly tripartite wage-policy dialogue and agreements since the NWC was formed in 1972. Second, the smooth cooperation between union and nonunion workers and the NWC's effective tripartite coordination result in relatively small wage drifts (wage increases beyond those agreed upon in the central negotiations), which are in sharp contrast to the large wage drifts in Europe. Third, serving endogenously as an integrated part of Singapore's macroeconomic management strategy, the NWC has reduced government reliance on exogenous instruments such as fiscal policy and other nonwage income policies, whereas many European governments normally approach interventions from outside the labor market.

48.2 Complementarity of Wages and Exchange Rates

This section presents a policy-game model of wages and exchange rates at their levels.⁵ For analytical simplicity, consider a composite good traded internationally under the purchasing power parity. Suppose that workers (employees) exert their influence in cooperation with the government and employers rather than through militancy. Wages are negotiated between firms and workers each period. The representative firm hires workers to produce output q according to a production function $q(L) = L^\phi$ ($0 < \phi < 1$). In wage negotiations, the right-to-manage model is used whereby workers bargain with employers for desired wages and employers choose employment at the negotiated wage level.⁶

Let W be the nominal wage, E the nominal exchange rate measured as units of the domestic currency per unit of a foreign currency, P^f the price of the tradable good in the foreign currency, L^* the level of employment demanded as a function of the real wage W/EP^f , i.e., $L^* = L^*(W/EP^f)$, and r the alternative source of income in real terms (unemployment compensation, for example) when the negotiating parties fail to agree upon W . Additionally, a constant-elasticity-of-substitution function $U(x) = x^{1-\gamma}/(1-\gamma)$ ($0 < \gamma < 1$) determines a representative worker's increasing and concave utility of the earned real income x . Denote the gain to the firm from agreeing to any given wage by $G_F(W; EP^f)$ and the similar gain to workers by $G_L(W; EP^f, r)$, respectively. The role of the NWC is to incorporate any exchange-rate policy signal into the wage settlement process and guide the two negotiating parties to choose a wage level to maximize the generalized Nash product, a weighted geometric average of the gains to workers and to firms:

$$\begin{aligned} & \{G_F(W; EP^f)\}^s \{G_L(W; EP^f, r)\}^{1-s} \\ & \equiv \left\{ EP^f q \left[L^* \left(\frac{W}{EP^f} \right) \right] - WL^* \left(\frac{W}{EP^f} \right) \right\}^s \left\{ L^* \left(\frac{W}{EP^f} \right) \left[U \left(\frac{W}{EP^f} \right) - U(r) \right] \right\}^{1-s} \end{aligned} \quad (1)$$

where s ($0 < s < 1$) is a weight reflecting the relative bargaining strength of workers. The variation of s traces out all the negotiated wages between the reservation level and the monopoly level in a Nash bargaining.

The first-order optimality condition determines the negotiated wage as

$$w = e + p^f + (1 - \phi)s + \frac{\phi}{1 - \gamma} + \ln r, \quad (2)$$

⁵For a longer and more detailed version of the model discussed in this section, see Wu (1999). The author gratefully acknowledges the permission granted by Blackwell Publishing in this regard.

⁶For the right-to-manage model, see Nickell and Andrews (1983) and Oswald (1985).

where the lower-case variables denote their logarithms and the last two parametric terms are the result of Taylor's approximation. Equation (2) gives the wage negotiator's reaction function, which predicts the unit elasticities of wages with respect to the exchange rate and to the foreign price level, respectively. Furthermore, the Nash bargaining wage is greater the larger the workers' bargaining power (s), the higher the productivity (ϕ), the greater the unemployment compensation (r), and the greater the elasticity of marginal utility (γ).

Since employment is determined by the firms' demand for labor at the negotiated level of real wage, aggregate output, Y , is a decreasing function of the real wage. Let L^A be aggregate employment and $F(L^A)$ aggregate output. Since the Cobb-Douglas production function determines aggregate output, it then follows that $Y = F(L^A(w - e - p^f)) \equiv h(w - e - p^f) = (w - e - p^f)^{\phi/(\phi - 1)}$, with $h' = F'L^A < 0$ and $h'' = F''L^A + F'L^A' > 0$. The monetary-fiscal authority has a loss function, Ω , which involves a cost associated with the inflation rate, $\pi \equiv De + Dp^f$ (D is the first-order difference operator), and the deviations of the current account from its target level, $\Theta > 0$ (see Wu 1999, for the detailed derivation of current account balance, CA).

$$\Omega = \frac{a}{2}\pi^2 - (CA - \Theta). \quad (3)$$

The current account surplus, unlike inflation, is favorable to the government so that a negative weight is attached to the second term in the loss function.⁷ Inflation costs rise at an increasing rate with the rate of inflation, and the coefficient $a > 0$ measures the authority's intolerance of inflation. The authority's problem is to choose the exchange rate to minimize the loss function (3). The associated first-order condition is

$$a(De + Dp^f) + \frac{1-c}{1+\psi}h'(w - e - p^f) + \frac{\psi}{1+\psi}\tau\lambda = 0, \quad (4)$$

where c is marginal propensity to consume with respect to changes in disposable income, λ is the weight for changes in the exchange rate in the balance payment account as opposed to changes in real foreign reserves ($0 < \lambda < 1$), and τ^1 is the sensitivity of exchange rate appreciation with respect to the balance of payments ($\tau^1 < 0$); in addition, $-1 < \psi \equiv \frac{\beta}{1-\lambda} \left[\frac{d}{\theta} - (\rho - r) \right] < 0$, where β is offset coefficient between domestic and foreign components of the monetary base ($0 < \beta < 1$), θ is the proportion of CPF liabilities invested in government securities, d is marginal propensity to consume with respect to changes in real private saving, r is the real interest rate on the government debt, and ρ is the real rate of return on the debt-financed government overseas investment. Eq. (4) implicitly determines the government's reaction function of the exchange rate to changes in the wage level, which in turn influences the wage that wage negotiators in the private sector demand. Therefore,

⁷For a similar formulation of the loss function, see Barro and Gordon (1983), and Agénor (1994).

the reaction function also indirectly conveys a signal of the government preference about the desired wage level with respect to the optimal exchange rate.

The wage-negotiators' reaction function (2), with the unemployment compensation parameter r being normalized to unity, and the government's reaction function (4) jointly determine the static equilibrium (e^*, w^*) . The corresponding dynamic system in the neighborhood of (e^*, w^*) is

$$\dot{e} = g_1(e, w) = -a(De + Dp^f) - \frac{1-c}{1+\psi} h'(w - e - p^f) - \frac{\psi}{1+\psi} \tau \lambda, \quad (5)$$

$$\dot{w} = g_2(e, w) = e - w + p^f + (1 - \phi)s + \frac{\phi}{1 - \gamma}, \quad (6)$$

where $\dot{e}$ and $\dot{w}$ are the time derivatives of e and w . The dynamic system of the exchange rate and wages is stable as long as inflation is so costly that a depreciation increases inflation costs more than strengthens competitiveness, that is, $a > \frac{1-c}{1+\psi} h''$.⁸

The empirical analysis with a vector error correction (VEC) model below demonstrates the robustness of the negative relationship between the exchange rate and wages obtained from comparative statics.

There are three variables: LULC, the logarithm of unit labor costs of all sectors; LNEER, the logarithm of the nominal effective exchange rate; and LIMP, the logarithm of the import price index compiled using the US dollar prices. The quarterly data are from the *International Financial Statistics*, ranging from 1980:1 to 1997:1. The augmented Dickey-Fuller test suggests that the three variables are all $I(1)$ sequences.⁹ The model is set with four-quarter lags by the conventional criteria, and the Johanson cointegration test suggests that there are exactly two cointegrating equations. Formally, after depressing the lagged difference terms, the estimated vector error correction model with four-period lags can be written as

$$DY_t = \alpha \beta' Y_{t-1} + \dots + \varepsilon_t, \quad (7)$$

where $DY_t = (DLULC_t, DLNEER_t, DLIMP_t)'$, $Y_{t-1} = (LULC_{t-1}, LNEER_{t-1}, LIMP_{t-1}, 1)'$, α is a 3×2 matrix of the speed-of-adjustment parameters estimated as $[\alpha_1 \alpha_2 \alpha_3]'$ with $\alpha_1 = (-0.10, -0.01)$, $\alpha_2 = (-0.03, -0.18)$, and $\alpha_3 = (0.09, 0.07)$, β is a 4×2 matrix of the normalized cointegrating vectors given by $[\beta_1 \beta_2]'$ with $\beta_1 = (1, 0, -1.05, 0.09)'$ and $\beta_2 = (0, 1, 1.06, -9.46)'$, and $\varepsilon_t = (\varepsilon_{tLULC}, \varepsilon_{tLNEER}, \varepsilon_{tLIMP})'$ is the vector of white-noise disturbances.

The estimated cointegrating coefficients in the matrix β are significant with wide margins even at 1% significance level. According to the two cointegration equations, responding to an increase of one percentage in import prices, wages increase by

⁸The mathematical results of dynamics as well as comparative statics are available from the author upon request.

⁹The augmented Dickey-Fuller values for LULC, LNEER, and LIMP are all below the 10% MacKinnon critical value in absolute terms.

1.05% and the exchange rate decreases by 1.06% in the long run. It follows that the purchasing power of wages in Singapore, measured in a basket of foreign currencies, has been rising in terms of imported goods. Derived from the two estimated cointegrating equations, the deterministic long-run equilibrium relationship can be described as

$$LULC = 9.28 - 0.99LNEER. \quad (8)$$

Equation (8) says that on average, each percentage of wage growth goes hand in hand with an approximately equal percentage of the Singapore currency appreciation vis-à-vis a basket of foreign currencies. The estimated speed-of-adjustment coefficients in α reflect the dynamic adjustment mechanism and support the robustness of the long-term equilibrium relationship. Suppose that one-unit positive shock in import prices results in a negative deviation in the unit labor cost and a positive deviation in the exchange rate from the previous period's stationary equilibrium, respectively. In response to the disequilibrium errors, the growth of unit labor cost increases by 10% as suggested by the first adjustment coefficient in α_1 , and the appreciation rate increases by 18% as suggested by the second adjustment coefficient in α_2 . Both the speed-of-adjustment coefficients are significant at 1% level and convergent as well.¹⁰

48.3 Policy Games in Wage Growth and Exchange Rate Appreciation

This section explicitly models the tripartism between employers, union workers, and the government as the institutional foundation to form the NWC objective function.¹¹ Employers as a whole concern themselves with the competitiveness of their products in the world market, which hinges highly upon relative unit labor cost. Union workers, on the other hand, are interested in maintaining a balance between employment and the growth of real income. Unlike the groups of union workers and employers, the government targets healthy macroeconomic performance characterized by a balance between inflation and unemployment.

The growth rate of $ULC(g_{ULC})$ is a weighted average of the wage-growth rate (g_w) and the inflation rate (π): $g_{ULC} = (1 - \theta)g_w + \theta\pi$, where the weight θ is actually

¹⁰With one standard-deviation innovation in import prices (LIMP) leading to a positive response of unit labor cost (LULC) and a negative response of the exchange rate (LNEER), both responses peak almost simultaneously at the fourteenth quarter after the shock; after that, both of them show a tendency to decay. Consistent with the cointegrating relationship discussed earlier, the pattern in which the wage response mirrors inversely the exchange rate response holds uniformly for all the possible orderings of the Choleski decomposition.

¹¹With kind permission of Springer Science and Business Media, the author has drawn on a longer version of Wu (2004) in the writings of this section and the next.

the parameter in a power function of labor productivity.¹² Denote the growth rate of foreign unit labor cost by g_{ULC^f} , then the expression $[(1 - \theta)g_w + \theta\pi + g_{NEER} - g_{ULC^f}]$ describes the evolution of relative unit labor costs. Formally, the NWC chooses the growth rate of wages to minimize its loss function

$$\begin{aligned} Loss_{NWC} = & \alpha_1 [(1 - \theta)g_w + \theta\pi + g_{NEER} - g_{ULC^f}] + \alpha_2 \left[\frac{1}{2} (U - \hat{U})^2 + \frac{\gamma}{2} (\pi - \hat{\pi})^2 \right] \\ & + \alpha_3 [\beta(g_w - \pi) + U] \end{aligned} \quad (9)$$

where $\hat{U}$ and $\hat{\pi}$ are the rates of unemployment and inflation targeted by the government, β is the union workers' loss weight of real income relative to unemployment, γ is the government's loss weight of inflation relative to unemployment, and α_1 , α_2 , and α_3 represent the three weights associated with the loss functions of employers, the government, and union workers, respectively (these α 's are the proxy parameters for the NWC participants' bargaining power). Note that $\gamma > 0$, $\beta < 0$, $\alpha_i > 0$, and $\sum \alpha_i = 1$. The first term in (9) describes the cost to employers of deteriorating relative unit labor cost, the second term represents the cost to the government when the unemployment rate and the inflation rate are off their targets, and the last term characterizes the cost to union workers when the real wage-growth rate falls or the unemployment rate rises.

The resulting optimal wage-growth rate responds to changes in economic conditions according to the following rule of reaction:

$$\begin{aligned} g_w = & \frac{A_1}{A_0} + \frac{A_2}{A_0} \pi_{Nop} + \frac{A_3}{A_0} g_{NEER} + \frac{A_4}{A_0} g_{w-2} + \frac{A_5}{A_0} \pi_{-2} + \frac{A_6}{A_0} \pi_{OP-2} + \frac{A_7}{A_0} g_{w-3} \\ & + \frac{A_8}{A_0} \pi_{-3} + \frac{A_9}{A_0} CPFC + \frac{A_{10}}{A_0} \hat{U} + \frac{A_{11}}{A_0} \hat{\pi} \end{aligned} \quad (10)$$

where $A_j (j = 0, 1, \dots, 11)$ are the functions of the structural parameters in the inflation equation, the unemployment-rate equation, and the unit-labor-cost growth equation, and the relative weights in the NWC's loss function (9). The values of these coefficient functions (A_j 's) are sensitive to the model's economic structure.

The other policy-game player, the MAS, manipulates the exchange rate to improve the trade-off between the imported inflation and the international competitiveness of Singapore's goods and services,¹³ which depends upon the real effective exchange rate, i.e., the relative unit labor cost in this chapter. Although the benefits of currency depreciation to the export sector can be lost to imported inflation and the resulting wage-price spiral that builds up in the medium-term horizon of 3 or more

¹²For the modeling and econometric specification of unit labor cost equation as well as price equation and unemployment equation, see the appendices in Wu (2004).

¹³See Teh and Shanmugaratnam (1992) and Carling (1995) for the analyses of Singapore's monetary policy via exchange-rate targeting.

years,¹⁴ maintaining a strong currency is detrimental to the export sector in the short run.

Let g_E be the actual real appreciation rate, which equals $(1 - \theta)g_w + \theta\pi + gNEER - gULC^f$, $\hat{g}$ the real appreciation rate targeted by the MAS, and δ the loss weight of the deviation of the inflation rate from its target relative to the deviation of real appreciation. The MAS selects the nominal appreciation rate, $gNEER$, to minimize its loss function:

$$Loss_{MAS} = \frac{1}{2}(g_E - \hat{g}_E)^2 + \frac{\delta}{2}(\pi - \hat{\pi})^2 \quad (11)$$

The first-order condition generates the MAS' rule of reaction:

$$g_{NEER} = \frac{B_1}{B_0}g_{ULC^f} + \frac{B_2}{B_0}g_w + \frac{B_3}{B_0}g_{w-2} + \frac{B_4}{B_0}\pi_{-2} + \frac{B_5}{B_0}\pi_{NOP} + \frac{B_6}{B_0}\pi_{OP-2} + \frac{B_7}{B_0}\hat{g}_E + \frac{B_8}{B_0}\hat{\pi} \quad (12)$$

where B_k 's ($k = 0, 1, \dots, 8$) are functions of the structural parameters and weights as A_j 's in (10).

In this model, the Nash game requires that a policy-making institution react to the optimal policy move made by the other policy-making institution as well as to the state of the economy. At the equilibrium, each institution's policy response is the best not only for the economy but also for the optimal policy of the other institution. Solving simultaneously the system of two non-Nash policy response functions, i.e., (10) and (12) with the estimated structural coefficients, produces the Nash equilibrium characterized by the Nash appreciation rate of NEER ($g * NEER$) and Nash wage-growth rate ($g * w$) (both in their implicit forms) below:

$$g * w = f(\pi_{NOP}, g_{ULC^f}, \pi_{OP-2}, g_{w-2}, \pi_{-2}, g_{w-3}, \pi_{-3}, CPFC, \hat{g}_E, \hat{U}, \hat{\pi}) \quad (13)$$

$$g * NEER = h(\pi_{NOP}, g_{ULC^f}, \pi_{OP-2}, g_{w-2}, \pi_{-2}, g_{w-3}, \pi_{-3}, CPFC, \hat{g}_E, \hat{U}, \hat{\pi}) \quad (14)$$

In contrast, the non-Nash game simply takes feedback from the state of the economy over a set of current and lagged state variables. Under this rule, any policy variable under the control of one institution does not react to a policy variable under the control of the other institution, that is, only the currently observed appreciation rate enters the NWC's reaction function, whereas only the currently observed wage growth rate enters the MAS's reaction function. By estimating the structural coefficients in the non-Nash policy response functions for the NWC and MAS, i.e., (10)

¹⁴ See Low (1994).

and (12), the non-Nash game can be reduced to one in which the policy sensitivity depends only on the weighting parameters and policy targets.¹⁵

The stability of Nash equilibrium depends on whether the recursive relations determined by (10) and (12) will yield a damped or an explosive time path of oscillation once the Nash equilibrium is disturbed. As shown in Wu (2004), the MAS response function (12) with estimated structural parameters is negatively sloped (for a reasonable value range of δ), and the similarly estimated NWC response function (10) is positively sloped, and then the stability condition for the Nash equilibrium requires that the NWC response function be flatter than the MAS response function in the policy space ($gNEER$, gW). This condition is not satisfied, however. It therefore follows that with the appropriate estimates of structural parameters the Nash equilibrium is not stable and it is more meaningful to concentrate on the non-Nash equilibrium.

48.4 Complementarity of Non-Nash Wage Growth and Exchange-Rate Appreciation

Fixing the policy targets and assigning different values to the relative weights α_t , β , and γ make it possible to simulate the computable time-paths of the non-Nash optimal appreciation rate, $gNEER$, and the non-Nash optimal wage growth rate, gW , over different economic scenarios. The purpose of simulation is to mimic non-Nash policy strategies and thus examine their sensitivity to the game-players' bargaining parameters and the policy stance.

There are three economic scenarios for simulation. In the benchmark case of Goldilocks economy (scenario 1), employees are equally concerned with real wage decline and unemployment ($\beta = -1$). The deviations from the government's targeted inflation rate are equally penalized as those from the targeted unemployment rate ($\gamma = 1$). And the MAS weights equally the deviations of the real exchange rate and inflation rate from their targeted levels ($\delta = 1$). In addition, the targeted rates of inflation and unemployment are respectively set at 2 and 3%, approximately, to reflect their long-term trend in the period; the targeted rate of real effective exchange-rate appreciation is chosen as 3% based on an 8-year moving average since 1988; and the targets specified above continue to apply to the other two economic scenarios. In a recession (scenario 2), the threat of recession prevents employees from demanding too much of real wage growth so that β falls in its absolute value ($\beta = -0.8$). The government and the monetary authority's inflation weights assume a smaller value compared with the Goldilocks economy ($\gamma = \delta = 0.8$). The third scenario concerns an inflationary economy in which the monetary and fiscal authority weighs inflation more than the targeted real competitiveness ($\delta = 1.2$). In the NWC's loss function, the government's inflation target now also takes a greater weight than the unemployment target ($\gamma = 1.2$), and meanwhile, the inflation threat naturally raises the employees' concern with their real income ($\beta = -1.2$).

¹⁵ See Wu (2004).

Table 1 Correlation between wage growth and NEER appreciation

Non-Nash game simulation				
	Actual	Scenario 1	Scenario 2	Scenario 3
Correlation Coefficient	0.644	0.588	0.646	0.520

How do the growth rate of wages and the appreciation rate of exchange rates work together in Singapore? Table 1 presents the correlation coefficients between the NEER appreciation rate and non-Nash wage growth rate in all the three simulated scenarios as well as the actually observed correlation coefficient.¹⁶ As in Table 1, all the simulation-based correlation coefficients are positive for the non-Nash regime. It follows that the two policies are complements in a non-Nash environment. Instead of responding optimally to each other, the non-Nash strategies work in such a way that at least one strategy acts independently without taking into consideration the intended target of the other. Hence, the two strategy variables tend to be relatively impartial in balancing and achieving their own targets. Furthermore, the observed positive correlation between actual wage growth and actual exchange-rate appreciation also matches the pattern for the simulated non-Nash outcome; it does so especially in scenario 2.¹⁷

48.5 Conclusion

Singapore government's commitment to and continuous participation in the annual tripartite collective bargaining over wage growth signifies the effectiveness of the NWC's adaptable stance and flexible wage policy in smoothing out business cycles, which detracts from the conventional wisdom on wage rigidity and its macroeconomic implications. This chapter explores the way Singapore policymakers deploy wage policy in coordination with its exchange-rate policy to achieve macroeconomic stability. The theoretical result from the Nash bargaining in the level of wages and exchange rates suggests that in the long run, wages increase one percentage point for about every percentage point appreciation in the exchange rate, which is well supported by the cointegration and error-correction analysis.

Furthermore, for the period studied, Singapore's tripartite collective bargaining (through NWC) in the growth rate of wages seems to have followed the non-Nash game practice as opposed to the Nash game, as the latter is unstable. A number of structural factors could have actually prevented the NWC from optimally reacting to

¹⁶For a given simulated scenario, the correlation coefficient does not vary with different bargaining cases because the parameters that reflect bargaining power are constant over time and they do not appear in coefficients of any variable in equations (14) and (15).

¹⁷Clearly, scenario 2 (recession) cannot characterize the 1987–1995 period in Singapore. The closest match of the simulated correlation with actual correlation in the scenario only suggests that there are some similarities between a recession period with low inflation or deflation and a high-growth period with low inflation; however, the simulation that is based almost exclusively on inflation-related parameters cannot distinguish one from the other.

the best move made by the MAS, such as asymmetry in the decision-making frequency (high frequency on the part of MAS vs. the low frequency on the part of NWC), asymmetric information between policy players as well as their overlapping interests, or simply any barrier in the institutional structure that makes a full-fledged interaction between policy players unrealistic. Both the non-Nash rule simulation and actual observations indicate that the Singapore dollar exchange rate appreciation has acted as a complement to wage growth. Indeed, Singapore currency has clearly exhibited a trend of appreciation vis-à-vis a basket of foreign currencies during economic upturns while the growth of labor earnings are rising and a trend of depreciation during economic downturns while the wage growth are declining.

Acknowledgments The author is grateful to Kenneth S. Chan, Michele Fratianni and two anonymous referees for their helpful comments when the parts of this work were published in *Pacific Economic Review* and *Open Economies Review*, respectively (see Wu, 1999 and Wu, 2004 in References).

References

- Abraham, F., K.D. Bruyne, and Isabel Van Der Auwera. 2000. Will wage policy succeed in euro-land? The case of Belgium. *Cahiers Economiques de Bruxelles* (168): 443–480.
- Agénor, P.R. 1994. Credibility and exchange rate management in developing countries. *Journal of Development Economics* 45 (1): 1–16.
- Barro, R.J., and D.B. Gordon. 1983. Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics* 12 (July): 101–121.
- Calmfors, L. 1998. Macroeconomic policy, wage setting, and employment – What difference does the EMU make? *Oxford Review of Economic Policy* 14 (3): 125–151.
- Carling, R.G. 1995. Fiscal and monetary policies. In *Singapore: A case study in rapid development*, IMF occasional paper 119, ed. K. Bercuson. Washington, DC: International Monetary Fund.
- Hahn, F.H. 1983. *Money and inflation*. Cambridge, MA: The MIT Press.
- Karadeloglou, P., C. Papazoglou, and G. Zombanakis. 2000. Exchange rate vs. supply-side policies as anti-inflationary devices: A comparative analysis for the case of Greece. *Archives of Economic History* 11: 47–61.
- Lawler, P. 2000. Union wage setting and exchange rate policy. *Economica* 67: 91–100.
- Low, V. 1994. Chap. 3 The MAS model: Structure and some policy simulations. In *Outlook for the Singapore economy*, 20–32. Singapore: Trans Global Publishing.
- Nickell, S.J., and M. Andrews. 1983. Unions, real wages and employment in Britain, 1951–79. *Oxford Economic Papers* 35 (Supplement): 183–206.
- Oswald, A.J. 1985. The economic theory of trade unions: An introductory survey. *Scandinavian Journal of Economics* 87 (2): 160–193.
- Otani, I., and C. Sassanpour. 1988. Financial, exchange rate, and wage policies in Singapore, 1979–86. *International Monetary Fund Staff Papers* 35 (3): 474–495.
- Perry, G.L. 1967. Wages and the guideposts. *American Economic Review* 57 (3): 897–904.
- Schultz, G.P., and R.Z. Aliber. 1966. *Guidelines: Formal controls and marketplace*. Chicago: University of Chicago Press.
- Teh, K.P., and T. Shanmugaratnam. 1992. Exchange rate policy: Philosophy and conduct over the past decade. In *Public policies in Singapore: Changes in the 1980s and future signposts*, ed. L. Low and M.H. Toh. Singapore: Times Academic Press.
- Wu, Y. 1999. Macroeconomic cooperation of exchange rate and wage movements under quasi-sterilization: Singapore experience. *Pacific Economic Review* 4 (2): 195–212.
- . 2004. Singapore's collective bargaining in a game of wage and exchange rate policies. *Open Economies Review* 15 (3): 273–289.



The Le Chatelier Principle of the Capital Market Equilibrium

49

Chin W. Yang, Ken Hung, Matthew Brigida, and John A. Fox

Abstract

This chapter purports to provide a theoretical underpinning for the problem of the Investment Company Act. The theory of the Le Chatelier principle is well known in thermodynamics: The system tends to adjust itself to a new equilibrium as far as possible. In capital market equilibrium, added constraints on portfolio investment in each stock can lead to inefficiency manifested in the right-shifting efficiency frontier. According to the empirical study, the potential loss can amount to millions of dollars coupled with a higher risk-free rate and greater transaction and information costs.

Keywords

Markowitz model · Efficient frontiers · With constraints · Without constraints · Le Chatelier principle · Thermodynamics · Capital market equilibrium · Diversified mutual funds · Quadratic programming · and Investment Company Act

John A. Fox: deceased.

C. W. Yang (✉)

Clarion University of Pennsylvania, Clarion, PA, USA

National Chung Cheng University, Chia-yi, Taiwan

e-mail: yang@clarion.edu

K. Hung

Division of International Banking and Finance Studies, Texas A&M International University, Laredo, TX, USA

M. Brigida

SUNY Polytechnic Institute, Utica, NY, USA

e-mail: matthew.brigida@sunypoly.edu

J. A. Fox

Matthew Brigida, Clarion University of Pennsylvania, Clarion, PA, USA

49.1 Introduction

In the wake of a growing trend of deregulation in various industries (e.g., utility, banking, and airline), it has become more and more important to study the responsiveness of the market to the exogenous perturbations as the system is gradually constrained. According to the law of thermodynamics, the system tends to adjust itself to a new equilibrium by counteracting the change as far as possible. This law, the Le Chatelier's principle, was applied to economics by Samuelson (1949, 1960, 1972), Silberberg (1971, 1974, 1978), and to a class of spatial equilibrium models: linear programming, fixed demand, quadratic programming, and full-fledged spatial equilibrium model by Labys and Yang (1996). Recently, it has been applied to optimal taxation by Diamond and Mirrlees (2002).

According to subchapter M of the Investment Company Act of 1940, a diversified mutual fund cannot have more than 5% of total assets invested in any single company and the acquisition of securities does not exceed 10% of the acquired company's value. By meeting this diversification threshold, funds are considered "pass-through" entities enabling capital gains and income taxes to accrue to the fund's investors. This diversification rule, on the one hand, reduces the portfolio risk according to the fundamental result of investment theory. On the other hand, more and more researchers begin to raise questions as to the potential inefficiency arising from the Investment Company Act (see Elton and Gruber, 1991; Roe, 1991; Francis, 1993; Kohn, 1994). Further, Almazan et al. (2004) documents inefficiencies from a broad set of mutual fund investment constraints. With the exception of the work by Cohen and Pogue (1967), Frost and Savarino (1988), and Loviscek and Yang (1997), there is very little evidence to refute or favor this conjecture.

Empirical findings (e.g., Loviscek and Yang, 1997) suggest that over 300 growth mutual funds evaluated by Value Line show that the average weight for the company given the greatest share of a fund's assets was 4.29%. However, the Le Chatelier's principle in terms of the Investment Company Act has not been scrutinized in the literature of finance. The objective of this chapter is to investigate the Le Chatelier principle applied to the capital market equilibrium in the framework of the Markowitz portfolio selection model.

49.2 The Le Chatelier Principle of the Markowitz Model

In a portfolio of n securities, Markowitz (1952, 1956, 1959, 1990, 1991) formulated the portfolio selection model in the form of a quadratic programming as shown below:

$$\min_{x_i} v = \sum_{i \in I} \sum_{j \in J} x_i x_j \sigma_{ij} \quad (1)$$

subject to

$$\sum_{i \in I} r_i x_i \geq k \quad (2)$$

$$\sum_{i \in I} x_i = 1 \quad (3)$$

$$x_i \geq 0 \forall i \in I \quad (4)$$

where

x_i = proportion of investment in security i .

σ_{ii} = variance of rate of return of security i .

σ_{ij} = covariance of rate of return of security i and j .

r_i = expected rate of return of security i .

k = minimum rate of return of the portfolio.

I and J are sets of positive integers.

The resulting Lagrange function is therefore

$$L = v + \lambda \left(k - \sum r_i x_{ij} \right) + \gamma \left(1 - \sum x_i \right) \quad (5)$$

The solution to the Markowitz is well known (1959). The Lagrange multiplier of constraint Eq. (48.2) assumes the usual economic interpretation: change in total risk in response to an infinitesimally small change in k while all other decision variables adjust to their new equilibrium levels, i.e., $\lambda = dv/dk$. Hence, the Lagrange multiplier is of utmost importance in determining the shape of the efficiency frontier curve in the capital market. Note that values of x_{is} are unbounded between 0 and 1 in the Markowitz model. However, in reality, the proportion of investment on each security many times cannot exceed a certain percentage to ensure adequate diversification. As the maximum investment proportion on each security decreases from 99% to 1%, the solution to the portfolio selection model becomes more constrained, i.e., the values of optimum x s are bounded within a narrower range as the constraint is tightened. Such impact on the objective function v is straightforward: As the system is gradually constrained, the limited freedom of optimum x s gives rise to a higher and higher risk level as k is increased. For example, if parameter k is increased gradually, the Le Chatelier principle implies that in the original Markowitz minimization system, isorisk contour has the smallest curvature to reflect the most efficient adjustment mechanism:

$$\left| \left(\frac{\partial^2 v}{\partial k^2} \right) \right| \leq \left| \left(\frac{\partial^2 v}{\partial k^2} \right) \right| \leq || \quad (6)$$

where v^* and v^{**} are the objective function (total portfolio risk) corresponding to the additional constrains of $x_i < s^*$ and $x_i < s^{**}$ for all i and $s^* > s^{**}$ represent different investment proportions allowed under V^* and V^{**} and abs denotes absolute value. Via the envelope theorem (Dixit, 1990) we have

$$\frac{d\{L(x_i(k), k) = v(x_i(k))\}}{dk} = \frac{\partial\{L(x_i(k), k) = v(x_i(k))\}}{\partial k} = \lambda \vee x_i = x_i(k) \quad (7)$$

hence, Eq. (48.6) can be rewritten as

$$\left| \left(\frac{\partial \lambda}{\partial k} \right) \right| \leq \left| \left(\frac{\partial \lambda}{\partial k} \right) \right| \leq \parallel \quad (8)$$

Equation (48.8) states that the Lagrange multiplier of the original Markowitz portfolio selection model is less sensitive to an infinitesimally small change in k than that of the model when the constraints are gradually tightened. Note that the Lagrange multiplier λ is the reciprocal of the slope of the efficiency frontier curve frequently drawn in investment textbooks. Hence, the original Markowitz model has the steepest slope for a given set of x_i s. However, the efficiency frontier curve of the Markowitz minimization system has a vertical segment corresponding to a range of low k s and a constant v . Only within this range do the values of optimum x s remain equal under various degrees of constraints. Within this range constraint Eq. (48.2) is not active, hence the Lagrange multiplier is zero. As a result equality relation holds for Eq. (48.8). Outside this range, the slopes of the efficiency frontier curve are different owing to the result of Eq. (48.8).

49.3 Simulation Results

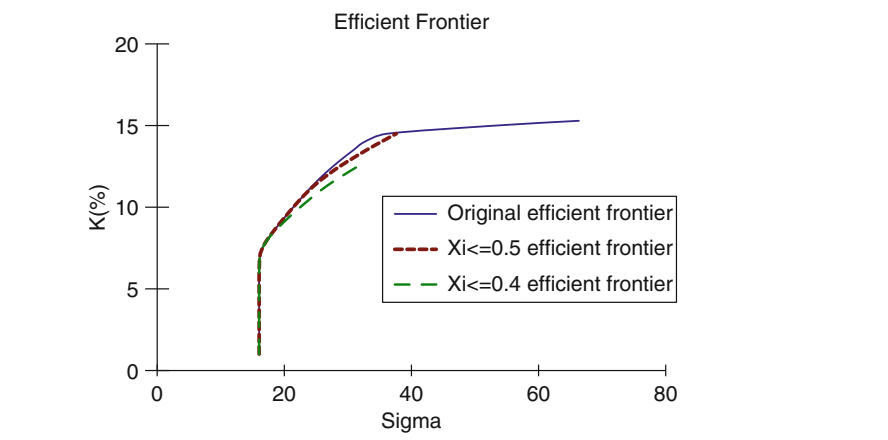
To verify the result implied by the Le Chatelier we employ a five-stock portfolio with $x_i < 50\%$ and $x_i < 40\%$. The numerical solutions are reported in Table 1. An examination of Table 1 indicates that the efficiency frontier curve is vertical and all optimum x s are identical between $0.001 < k < 0.075$. After that, the solutions of x s begin to change for the three models. Note that the maximum possible value for x_4 remains 0.4 throughout the simulation for $k > 0.075$ for the model with the tightest constraint $x_i < 0.4$. In the case of $x_i < 0.5$, a relatively loosely constrained Markowitz system, all the optimum values of decision variables remain the same as the original Markowitz model between $0.01 < k < 0.1$. Beyond that range, the maximum value of x_4 is limited to 0.5. As can be seen from Table 1, the total risk v responds less volatile to the change in k in the original unconstrained Markowitz system than that in the constrained systems. In other words, the original Markowitz minimization system guarantees a smallest possible total risk due to the result of the Le Chatelier's principle: A thermodynamic system (risk-return space) can most effectively adjust itself to the parametric change (temperature or minimum rate of return of a portfolio or k) if it is least constrained.

49.4 Policy Implications of the Le Chatelier's Principle

As shown in the previous section, the efficiency frontier curve branches off to the right first for the most binding constraint of $x_i < s^{**}$. Consequently, the tangency point between the efficiency frontier curve and a risk-free rate on the vertical axis

Table 1 With a graph

LEAST CONSTRAINED SOLUTION (Original Markowitz Model)							SOLUTION WITH $x_i \leq 0.5$							SOLUTION WITH $x_i \leq 0.4$						
K(%)	$v(10^{-5})$	$x_1\%$	$x_2\%$	$x_3\%$	$x_4\%$	$x_5\%$	$v(10^{-5})$	$x_1\%$	$x_2\%$	$x_3\%$	$x_4\%$	$x_5\%$	$v(10^{-5})$	$x_1\%$	$x_2\%$	$x_3\%$	$x_4\%$	$x_5\%$		
1	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0		
2	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0		
3	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0		
4	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0		
5	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0		
6	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0	257.2	39.19	0	31.87	28.94	0		
7	260.8	35.02	0	32.6	32.38	0	260.8	35.02	0	32.6	32.38	0	260.8	35.02	0	32.6	32.38	0		
7.5	274.8	30.54	0	32.77	36.69	0	274.8	30.54	0	32.77	36.69	0	274.8	30.54	0	32.77	36.69	0		
8	299.3	25.82	0	33.27	40.91	0	299.3	25.82	0	33.27	40.91	0	300.5	24.91	0	34.55	40	5.39		
8.5	333.1	21.65	0	33.26	43.63	1.45	333.1	21.65	0	33.26	43.63	1.45	340.2	20.42	0	35.34	40	4.24		
9	371.2	17.82	0	32.92	45.73	3.53	371.2	17.82	0	32.92	45.73	3.53	387.7	15.93	0	36.13	40	7.94		
9.5	413.2	14.05	0	32.53	47.64	5.79	413.2	14.05	0	32.53	47.64	5.79	443	11.44	0	36.92	40	11.64		
10	459	9.68	0.58	32.17	49.59	7.98	459	9.68	0.58	32.17	49.59	7.98	506.2	6.95	0	37.71	40	15.34		
10.5	508.3	4.83	1.96	31.44	51.56	10.2	509.5	4.25	2.1	32.23	50	11.42	576.7	1.23	1.93	37.7	40	19.15		
11	560.9	0	3.53	30.46	53.55	12.46	567.5	0	2.66	32.03	50	15.31	656.5	0	0.21	36.45	40	23.34		
11.5	619.9	0	1.34	27.91	55.8	14.95	637.4	0	0	30.39	50	19.62	751.7	0	0	31.79	40	28.22		
12	687.5	0	0	24.31	58.11	17.58	724.5	0	0	25.39	50	24.62	866.3	0	0	26.79	40	33.22		
12.5	765.4	0	0	19.02	60.68	20.3	826.7	0	0	20.52	50	29.48	995.2	0	0	21.91	40	38.09		
13	854.3	0	0	13.73	63.2	23.07	949.7	0	0	15.53	50	34.48								
13.5	954	0	0	8.45	65.72	25.83	1086.8	0	0	10.65	50	39.45								
14	1064.6	0	0	3.16	68.25	28.59	1243.3	0	0	5.73	50	44.28								
14.5	1309.1	0	0	0	55.63	44.37	1417.7	0	0	0.79	50	49.21								
15	2847.3	0	0	0	20	80														
15.29	4402	0	0	0	0	100														



must occur at a higher risk-free rate. As the value of maximum investment proportion for each stock s decreases, i.e., the constraint becomes more binding, there is a tendency for the risk-free rate to be higher in order to sustain an equilibrium (tangency) state. Second, one can assume the existence of a family of isowelfare functions (or indifference curves) in the v - k space. The direct impact of the Le

Chatelier principle on the capital market equilibrium is a lower level of welfare measure due to the right branching off of the efficiency frontier curve. In sum, as the constraint on the maximum investment proportion is tighter, the risk-free rate will be higher and investors in the capital market will in general experience a lower welfare level. In particular, the 5% rule carries a substantial cost in terms of shifting of the efficiency frontier to the right. The study by Loviscek and Yang (1997) based on a 36-security portfolio indicates the loss is about 1–2 percentage points and the portfolio risk is 20–60% higher. Given the astronomical size of a mutual fund, 1–2 percentage point translates into millions of dollars potential loss in daily return. Furthermore, overdiversification would incur greater transaction and information cost, which speaks against the Investment Company Rule.

49.5 Concluding Remarks

In this note, we apply the Le Chatelier principle in the thermodynamics to the Markowitz's portfolio selection model. The analogy is clear: As a thermodynamic system (or the capital market in the v - k space) undergoes some parametric changes (temperature or minimum portfolio rate of change k), the system will adjust most effectively if it is least constrained. The simulation shows that as the constraint becomes more and more tightened, the optimum investment proportions are less and less sensitive. Via the envelop theorem, it is shown that investors will be experiencing a higher risk-free rate and a lower welfare level in the capital market, if a majority of investors in the capital market experience the same constraint, i.e., maximum investment proportion on each security. Moreover, the potential loss in daily returns can easily be in millions on top of much greater transaction and information costs.

References

- Almazan, Andres, Keith C. Brown, Murray Carlson, and David A. Chapman. 2004. Why constrain your mutual fund manager? *Journal of Financial Economics* 73: 289–321.
- Cohen, Kalman J., and Jerry A. Pogue. 1967. An empirical evaluation of alternative portfolio selection models. *Journal of Business* 40 (April): 166–189.
- Diamond, Peter, and Mirrlees, James A. 2002. Optimal taxation and the Le Chatelier principle. MIT Dept of Economics Working paper Series, September 24, 2002.
- Dixit, Avinash K. 1990. *Optimization in economic theory*. 2nd ed. Oxford University Press.
- Elton, Edwin J., and Martin J. Gruber. 1991. *Modern portfolio theory and investment analysis*. New York: John Wiley.
- Francis, Jack C. 1993. *Management of investment*. New York: McGraw-Hill.
- Frost, Peter A., and James E. Savarino. 1988. For Better Performance, Constrain Portfolio Weights. *Journal of Portfolio Management* 14 (Fall): 29–34.
- Kohn, Meir G. 1994. *Financial Institutions and Markets*. New York: McGraw-Hill.
- Labys, Walter C., and Chin-Wei Yang. 1996. Le Chatelier principle and the flow sensitivity of spatial commodity models. In *Recent advances in spatial equilibrium modeling: Methodology and applications*, ed. J.C.J.M. van de Bergh, P. Nijkamp, and P. Rietveld, 96–110. Berlin: Springer.

- Loviscek, Anthony L., and Chin-Wei Yang. Spring 1997. Assessing the impact of the investment company Act's diversification rule: A portfolio approach. *Midwest Review of Finance and Insurance* 11 (1): 57–66.
- Markowitz, Harry M. 1952. Portfolio selection. *The Journal of Finance* VII (1): 77–91.
- . 1956. The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly* 3: 111–133.
- . 1959. *Portfolio selection*. New York: John Wiley and Sons.
- . 1990. *Mean-variance analysis in portfolio choice and capital markets*. Basil: Blackwell.
- . 1991. Foundation of portfolio theory. *Journal of Finance* XLVI (2): 469–477.
- Roe, Mark J. 1991. Political element in the creation of a mutual fund industry. *University of Pennsylvania Law Review* 1469: 1–87.
- Samuelson, Paul A. 1949. *The Le Chatelier principle in linear programming*. Rand Corporation Monograph (August 1949).
- . 1960. An extension of the Le Chatelier principle. *Econometrica* (April): 368–379.
- . 1972. Maximum principles in analytical economics. *American Economic Review* 62 (3): 249–262.
- Silberberg, Eugene. June 1971. The Le Chatelier principle as a corollary to a generalized envelope theorem. *Journal of Economic Theory* 3: 146–155.
- . February 1974. A revision of comparative statics methodology in economics, or how to do economics on the back of an envelope. *Journal of Economic Theory* 7: 159–172.
- . 1978. *The structure of economics: A mathematical analysis*. New York: McGraw-Hill Book Company; *Value line, mutual fund survey*. New York: Value Line, 1994).



C. H. Ted Hong and Wen-Ching Wang

Abstract

This chapter not only provides a comparison of recent models in the valuation of mortgage-backed securities but also proposes an integrated model that addresses important issues of path-dependence, exogenous prepayment, transaction costs, mortgagors' heterogeneity, and the housing devaluation effect.

Recent research can be categorized into two frameworks: empirical and theoretical option pricing. Purely empirically derived models often consider estimation of the prepayment model and pricing of the mortgage-backed security as distinct problems, and thus preclude explanation and prediction for the price behavior of the security. Some earlier theoretical models regard mortgage-backed securities as default-free callable bonds, prohibiting the mortgagors from exercising the default (put) option, and therefore induce bias on the pricing of mortgage-backed securities. Other earlier models assume homogeneity of mortgagors and consequently fail to address important issues of premium burnout effect and the path-dependence problem.

The model proposed is a two-factor model in which the housing price process is incorporated to account for the effect of mortgagor's default and to capture the impact of housing devaluation. Default is correctly modeled in terms of its actual payoff through a guarantee to the investors of the security such that the discrepancy is eliminated by assuming mortgage securities as either default-free or uninsured. Housing prices have been rising at unsustainable rates nation wide, especially along the coasts, suggesting a possible substantial weakening in house appreciation at some point in the future. The effect of housing devaluation is specifically modeled by considering the possibility that the mortgagor might be

C. H. Ted Hong (✉)
Beyondbond, Inc, New York, NY, USA
e-mail: ted@beyondbond.com

W.-C. Wang
Robeco Investment Management, New York, NY, USA

restrained from prepayment even if interest rates make it advantageous to refinance.

Mortgagors' heterogeneity and the separation of exogenous and endogenous prepayments are explicitly handled in the model. Heterogeneity is incorporated by introducing heterogeneous refinancing transaction costs. The inclusion of heterogeneous transaction costs not only captures premium burn-out effect but also solves the path-dependence problem. Finally, the model separates exogenous prepayment from endogenous prepayment, and estimates their distinct magnitudes from observed prepayment data. This construction provides a better understanding for these two important components of prepayment behavior. The generalized method of moments is proposed and can be employed to produce appropriate parameter estimates.

Keywords

Devaluation trap · Exogenous and Endogenous prepayments · Generalized method of moments · Heterogeneity · Housing devaluation effect · MBS valuation · Option pricing theory · Path dependency · Premium burnout effect · Transaction costs of refinancing and default

50.1 Introduction

The main objective of this chapter is to gain a better understanding of the valuation of mortgage-backed securities. Mortgage-backed securities have attracted unprecedented investor interest over the last decade, spurring tremendous growth in the market for this important financial instrument. There are over \$7.7 trillion worth of residential mortgage loans outstanding, an amount far exceeding the size of the corporate debt market. Approximately \$5.1 trillion worth of securitized mortgage-backed securities and CMOs are outstanding, and well over \$1.8 trillion new mortgage-backed securities and whole loans pools are issued each year for the past 3 years.¹ Mortgage-backed securities are extensively held by every class of institutional investor, including commercial banks, saving institutions, insurance companies, mutual funds, and pension plans.

An in-depth study of the valuation of mortgage-backed securities is of interest to financial economists because mortgage-backed securities have unique characteristics that are distinct from other contingent claims, such as monthly amortization, negative convexity, premium burnout, and path-dependence. This chapter examines recent developments in the area of valuing mortgage-backed securities and proposes a model that accommodates these factors affecting the price of mortgage-backed securities.

¹Taken from the Federal Reserve Bulletin, Inside MBS & ABS, and UBS.

The core issue in valuing mortgage-backed securities is the modeling of the prepayment behavior of mortgagors in the pool backing the security. Continuous-time option pricing methodology has been a popular method in the mortgage-backed securities valuation because of the obvious parallel between the call option and the right of a mortgagor to prepay. In order to model the mortgagors' prepayment behavior more realistically, recent theoretical models have added modifications to the original stock option pricing theory framework. The first of these modifications broadly accounts for prepayment due to reasons exogenous to financial consideration, such as moving and job changes. The second group of modification addresses transaction costs. The third considers heterogeneity among mortgagors, and the fourth group discusses the separation of exogenous prepayment and endogenous prepayment.

The observation that homeowners clearly do not prepay as objectively as option pricing models imply has motivated many researchers to add prepayment functions that allow prepayments for reasons that are exogenous to purely financial considerations. Such research includes the work of Dunn and McConnell (1981a, b), and Brennan and Schwartz (1985), and most of the prepayment functions have been arbitrary. The main drawback of adding an arbitrary prepayment function is that it does not aid in the identification of the factors responsible for prepayment behavior. Identifying these factors would go a long way toward enhancing the explanatory power of the model.

Applying the option pricing theory to the valuation of residential mortgage-backed securities, one can see a departure from the perfect market assumption when homeowners face transaction costs upon refinancing or defaulting. For this reason, Dunn and Spatt (1986) and Timmis (1985) add homogenous refinancing transaction costs in their models to adjust the prepayment speeds from those implied in the frictionless economic environment. Kau et al. (1993a) also add the transaction cost of default in their modeling of the probability of default for residential mortgages.

Addressing mortgagors' heterogeneity is a more complex matter. Many earlier models assumed homogeneity among mortgagors to avoid complexity in the pricing process. However, the assumption of mortgagors' homogeneity fails to address the issue of premium burnout which is an important empirical effect of homeowner heterogeneity. And this assumption also results in a path-dependent problem when numerically solving the optimal refinancing strategies backwards. The premium burnout effect is the tendency of prepayments from premium pools to slow down over time, with all else held constant. If a large number of mortgagors have already prepaid, those remaining are likely to have a relatively low probability of prepaying. Conversely, the smaller the number of previous prepayments, the higher the probability of prepaying by the remaining mortgagors. The aforementioned path-dependent problem occurs because any mortgage pool contains a group of mortgagors who behave differently in their prepayment decisions: these mortgagors differ in their willingness or ability to prepay their loans under favorable circumstances. As a result, without knowing either the type of mortgagor or the entire path of interest rates from origination, backward optimization is not applicable because there is no way of knowing whether the earlier prepayment exercise is optimal.

Johnson and Van Drunen (1988), and Davidson et al. (1988) improve on the homogenous transaction cost model by introducing heterogeneous transaction models. They assume that different homeowners face different levels of refinancing transaction costs. In addition to the ability to capture the premium burnout, the inclusion of heterogeneous transaction costs also solves the path-dependent problem encountered when pooling individual mortgagors, who behave differently in their prepayment decisions.

Another common problem in existing models is the lack of differentiation between exogenous prepayment and endogenous prepayment. This lack of distinction between the two thereby precludes explanation of the interrelation between these important behavioral components. Endogenous prepayment refers to any prepayment decision that occurs in response to changes in underlying economic processes, such as the interest rate. Stanton (1990) incorporates an endogenous decision parameter that enables separate estimations of endogenous prepayment and prepayment for exogenous reasons. As a result, the explanatory power of the model is improved. In addition to the inclusion of the previously discussed modifications, our model introduces two adjustments. One is the treatment of mortgagors' right to default in the content of mortgage-backed securities valuation. And the other is the impact of the housing prices on prepayment behavior.

Although default has been modeled as a put option in the models of residential mortgages or commercial mortgage-backed securities, many earlier models have not incorporated it in the valuation of residential mortgage-backed securities. This is because government agency guarantees lead to the perception that securities are default-free. Default should be taken into consideration because there is a payoff difference between a guaranteed mortgage-backed security and a default-free security. The payoff from a guarantee in the event of default is the par amount rather than the market value of the security, thus producing an asymmetric return for investors.

In modeling default, we expand previous default-free models into a default-risky model in which the housing price process is included as a second-state variable. Default is explicitly modeled in terms of its actual payoff through a guarantee to the investors of the residential mortgage-backed security. This is in contrast to models for individual mortgages or commercial mortgages in which mortgages are neither insured nor guaranteed. Consequently, the payoff in the event of default in these cases is the value of the house. By correctly modeling the effect of default, our model reduces the discrepancy from assuming mortgage-backed securities as either default-free or uninsured.

The housing price process is incorporated in the model not only to account for the effect of default on security price, but also to determine its impact on the prepayment behavior of mortgagors. The effect of housing prices on prepayment is specifically modeled by considering the possibility that the mortgagor might be restrained from prepaying even if interest rates make it advantageous to refinance. This is because housing prices have fallen to the extent that the mortgagor is no longer qualified for refinancing.

The model we propose not only captures the fundamental characteristics of the mortgagors' prepayment behavior but it also combines parametric heterogeneity and

variability of the decision parameter to the extent that our model can come closer than previous models in describing empirical prepayment behaviors.

50.2 The Model

The central issue in valuing mortgage-backed securities is the treatment of prepayment uncertainty. The valuation model of mortgage-backed securities proposed here is based on the continuous-time option pricing methodology. This methodology treats the right of a mortgagor to prepay as a call option and the right to default as a put option. Modifications to the assumption of perfect capital markets and the principle that borrowers act to minimize the market cost of their mortgages are required to portray mortgagors' actual prepayment behavior in a more realistic manner.

According to Dunn and McConnell (1981a, b) and Brennan and Schwartz (1985), we allow mortgagors to prepay for reasons exogenous to purely financial considerations. In contrast to their models that assume arbitrary exogenous prepayment functions, our model utilizes the proportional hazard function and can be estimated from observable prepayment data.

To account for the fact that homeowners face transaction costs when they prepay or default on their mortgages, we follow Johnson and Van Drunen (1988). Consequently, we add heterogeneous refinancing transaction costs in our models to adjust the prepayment speeds from those implied in the frictionless economic environment. Following Kau et al. (1993a), we also add the transaction cost of default in modeling the effect of default.

Default has been modeled as a put option in the valuation of residential mortgages or commercial mortgage-backed securities. However, many models have not incorporated default in the valuation of residential mortgage-backed securities because government agency guarantees lead to the perception that securities are default-free. Moreover, there is a significant difference between the payoff of a guaranteed mortgage-backed security and that of a default-free security. The payoff from insurance in the event of default is the par amount rather than the market value of the security, producing an asymmetric return for investors.

Kau et al. (1992) develop a two-factor model for both prepayment and default only in the context of evaluating individual mortgages, where mortgages are considered as uninsured. The payoff from uninsured mortgages is the value of the house when the mortgage is defaulted. In our model, the payoff to the investor from default is explicitly modeled as insured mortgages. This eliminates the potential bias in the pricing of mortgage-backed securities.

A significant relationship between observed prepayment and housing prices data pointed out by Richard (1991) leads us a final adjustment of the two-factor model. The housing price process is brought in not only to account for the effect of default on security price, but also to determine its restraining effect on mortgagors' refinancing decisions.

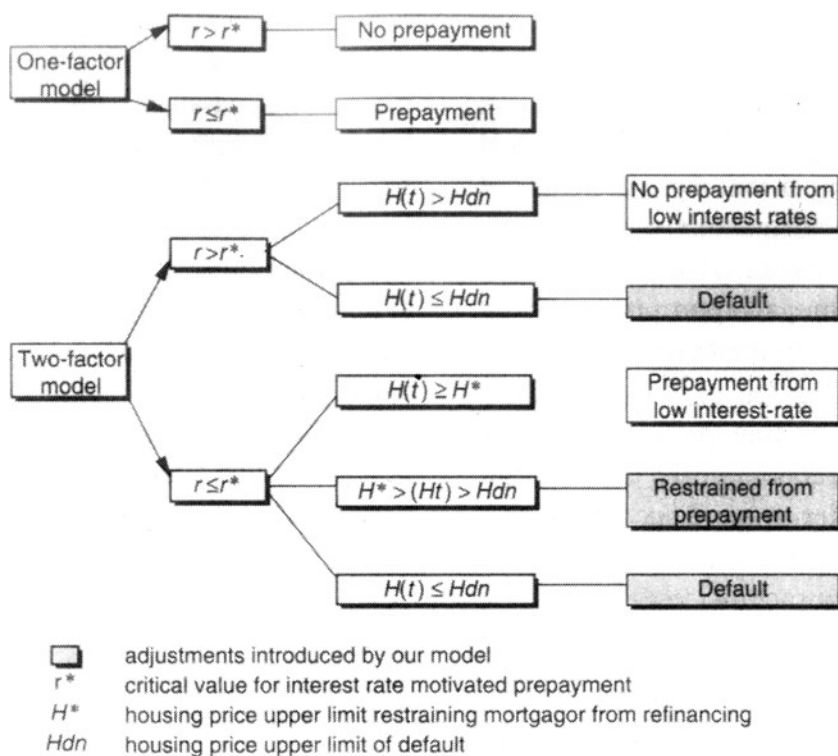


Fig. 1 Model trees

Figure 1 outlines these differences between one-and two-factor models and the innovations presented in this study.

In the one-factor model, the prepayment decision responds to the level of interest rates. The two-factor model adds two additional termination outcomes that follow from the level of housing prices. At very low housing prices, the mortgagors may default regardless of the interest rate in order to cut their losses. Finally, the mortgagor might be restrained from prepaying even if interest rates make it advantageous to refinance. This occurs when the housing prices fall to the extent that the new loan cannot cover the costs of refinancing.

In addition to capturing these fundamental characteristics of the mortgagor's termination behavior, this model aggregates the underlying pool of mortgages according to the heterogeneity of transaction costs. And it is the specification of heterogeneous transaction costs that also solves the path-dependent problem displayed by pooled mortgages.

The following first section pertains to the modeling of termination decisions affected by exogenous and endogenous factors, housing prices, and transaction costs. The later section introduces our model, which is a two-factor pricing

framework that provides exact security prices given underlying interest rate and housing prices processes, and precludes arbitrage opportunities.

50.2.1 Modeling Issues

50.2.1.1 Exogenous Prepayment

In practice, exogenous reasons for termination include factors such as relocation, death, divorce, or natural disasters. Exogenous prepayments are also known as turnover prepayments. A hazard function is used to model exogenous prepayment as follows:

$$\pi(t) = \lim_{\delta t \rightarrow 0^+} \frac{\Pr(\text{Exogeneous prepayment in } (t, t + \delta t) | \text{No Prepayment prior to } t)}{\delta t} \quad (1)$$

There are numerous parametric methods used in the analysis of duration data and in the modeling of aging or failure processes. We use the exponential distribution in the model for its simplicity. The distribution is characterized by the constant hazard function

$$\pi(t) = \pi, t \geq 0 \text{ and } \pi > 0. \quad (2)$$

The probability that an individual has not prepaid for exogenous reasons until time t is given by the survival function $S(t)$,

$$S(t) = e^{-\pi(t)} = e^{-\pi t}, \quad t \geq 0 \quad (3)$$

50.2.1.2 Endogenous Termination

A mortgage is terminated when mortgagors either prepay or default on their mortgages. Any termination which affects the cash flows passed through to the investors will have an impact on the price of the mortgage-backed securities. Throughout the model, endogenous termination is defined as any rational termination decision that occurs in response to underlying economic processes rather than personal considerations.

We assume that mortgagors maximize their current wealth, or equivalently, minimize their liabilities. Mortgagors' liabilities can be thought of as composed of three parts. The first part consists of owing the scheduled streams of cash flows associated with the mortgage. The second part constitutes their option to prepay at any time, which is equivalent to possessing a call option. And the third part consists of mortgagor's option to default, which functions as a put option. Option pricing theory is, therefore, an appropriate method for determining the value of mortgagors' mortgage liability.

A model of mortgage pricing should incorporate both refinancing transaction costs and default transaction costs in order to more accurately portray the decision-making processes of mortgagors. Although including transaction costs causes the resulting termination strategy to deviate from the perfect market assumption, the strategy still remains rational.

In order to derive the magnitude of endogenously determined termination, we follow Stanton (1990) and introduce ρ , which measures the frequency of mortgagors' termination decisions. The time between successive decision points is described as an exponential distribution. If we let T_i be one such decision point, and T_{i+1} the next, then

$$\Pr(T_{i+1} - T_i > t) = e^{-\rho t} \quad (4)$$

If mortgagors are continually re-evaluating their decisions, then the parameter ρ takes on a value of infinity. If mortgagors never make endogenous termination decisions and only terminate for exogenous reasons, then ρ takes on a value of zero. If ρ takes on a value between these limits, then this signifies that decisions are made at discrete times, separated on average by $1/\rho$.

Given this specification, the magnitude of endogenized termination can be estimated and studied. The contribution of this device is to separate the magnitude of endogenized termination from that of exogenous termination. It also serves to help understand the actual termination behavior of mortgagors. Without this specification, it would be difficult to know the proportion of termination from endogenous optimization decisions and the proportion due to exogenous factors.

Utilizing the definitions from Sects. 50.2.1.1 and 50.2.1.2, we notice that the optimal exercise strategy immediately leads to a statistical representation of the time to terminate for a single mortgagor. If termination is due exclusively to exogenous factors, then the termination rate is π and the survival function is defined as in Eq. 4. When termination occurs for endogenous reasons, the probability that the mortgagor terminates in a small time interval, δt , is the probability that the mortgagor neither prepays for exogenous reasons nor makes a rational exercise decision during this period. This survival function can be approximated by

$$S(t) = \begin{cases} e^{-\pi\delta t} \cdot e^{-\rho\delta t} = e^{-(\pi+\rho)\delta t} & \text{if endogenous termination} \\ e^{-\pi\delta t} & \text{if no endogenous termination} \end{cases} \quad (5)$$

50.2.1.3 Transaction Costs and Aggregation of Heterogeneous Mortgages

The cash flows that accrue to the investor of a mortgage-backed security are not determined by the termination behavior of a single mortgagor, but by that of many mortgagors within a pool. To cope with the path-dependent problem caused by the heterogeneity within a pool of mortgages, we assume that the different refinancing transaction costs each mortgagor faces is the only source of heterogeneity. Although the costs of initiating a loan vary among different types of mortgages, some of the

most common costs borrowers face include credit report, appraisal, survey charges, title and recording fees, proration of taxes or assessments, hazard insurance, and discount points.

The transaction costs of individual mortgagors are drawn from a univariate discrete distribution, which allows for underlying heterogeneity in the valuation of the mortgage-backed security. A better way to choose the underlying distribution that represents this heterogeneity would be to look at summary statistics of transaction costs actually incurred by mortgagors when they refinanced. A discrete rectangular distribution is chosen for its simplicity and the task of determining which distribution improves the fit is left for future research.

The value of the security is equal to the expected value of the pool of mortgages weighted by the proportions of different refinancing transaction cost categories. Suppose that each X_i (the refinancing transaction costs faced by mortgagor i) is drawn from a discrete rectangular, or uniform distribution

$$\Pr(x = a + ih) = M^{-1}, i = 1, \dots, M \quad (6)$$

Various standard forms are in use. For this application, we set $a = 0$, $h = RM^{-1}$, so that the values taken by x are RM^{-1} , $2RM^{-1}$, $\dots$, R . The upper bound R of the transaction cost is set at 10%. The distribution for the transaction costs is then defined as:

$$\Pr\left(x = i \frac{0.1}{M}\right) = M^{-1}, i = 1, \dots, M \quad (7)$$

In principle, given any initial distribution of transaction costs, it is possible to value a mortgage-backed security backed by a heterogeneous pool of mortgages in a manner similar to the valuation of a single mortgage. If the value of individual mortgages is known, then the value of the pool is the sum of these individual values. When the value of individual mortgages is not known, but a distribution of transaction costs is generated that accounts for heterogeneity, the expected value of a pool of mortgages is the sum of the transaction cost groups times the probability of their occurrence in the pool.

Recall from Sect. 50.2.1.2 that for a given transaction cost X_i and state of the world, if any mortgagor finds it optimal to terminate, the hazard rate is the sum of the exogenous prepayment rate, π , and the endogenized termination rate, ρ . If it is not optimal to terminate, the hazard rate falls back to the background exogenous prepayment rate π .

Models that neither permit the estimation of ρ nor consider exogenous factors in the prepayment decision imply that $\rho = \infty$ and $\pi = 0$, and the single-transaction cost level predicts that all mortgages will prepay simultaneously. Adding heterogeneous transaction costs addresses the problem of path dependence, however, keeping the same parameter values still does not permit hesitation in the prepayment decision. Although prepayment rates fluctuate, in reality, they do tend to move fairly smoothly. The effect of setting ρ to a value other than ∞ is to permit a delay even when it is optimal to prepay. And prepayment need not occur at all if interest rates or housing

prices change such that it is no longer optimal. The actual value of ρ determines how fast this drop occurs. Thus, combining parametric heterogeneity and variability of the parameter ρ would allow the model to come closer than previous rational models to describe empirical prepayment behavior.

50.2.2 A Model for Pricing Mortgage-Backed Securities

50.2.2.1 Termination Decision of a Single Mortgagor

The following is a model of rational prepayment behavior of mortgages that extends the rational prepayment models of Stanton (1990) and Kau et al. (1993b). Mortgagors may terminate their mortgages for endogenous financial reasons that include interest rates and housing prices, or for exogenous reasons. They also face transaction costs, which are used to differentiate mortgagors and solve the path-dependent problem. Mortgagors choose the strategy that minimizes the market value of the mortgage liability.

The following assumptions are employed:

1. Trading takes place continuously and there are no taxes or informational asymmetries.
2. The term structure is fully specified by the instantaneous riskless rate $r(t)$. Its dynamics are given by.

$$dr = \kappa(\mu_r - r)dt + \sigma_r \sqrt{r} dz_r \quad (8)$$

3. The process to capture the housing price is assumed to follow a Constant Elasticity of Variance (CEV) diffusion process

$$dH = \mu_H H dt + \sigma_H H^{\gamma/2} dz_H \quad (9)$$

where $\mu_H, \sigma_H > 0$, $0 < \gamma < 2$, and $\{z_H(t), t \geq 0\}$ is a standard Wiener Process, which may be correlated with the process $\{z_r(t), t \geq 0\}$. When $\gamma = 2$, the process is lognormal.

The underlying state variables in the model are the interest rate $r(t)$ and the housing price $H(t)$. By applying the arbitrage argument, the value of the i th mortgage liability $V^i(r, H, t)$ satisfies the following partial differential equation:

$$\begin{aligned} & \frac{1}{2} \sigma_r^2 r V_{rr}^i + \rho \sigma_r \sigma_H \sqrt{r} H^{\gamma/2} V_{rH}^i \\ & + \frac{1}{2} \sigma_H^2 H^\gamma V_{HH}^i + [\kappa(\mu_r - r) - \lambda r] V_r^i \\ & + r H V_H^i + V_t^i - r V^i = 0 \end{aligned} \quad (10)$$

where λr represents factor risk premium.

The value of the mortgage liability is also required to satisfy the following boundary conditions:

1. At maturity T , the value of a monthly amortization bond is equal to the monthly payment:

$$V^i(r, H, T) = MP$$

2. As r approaches infinity, the payoff of the underlying mortgage bond approaches zero:

$$\lim_{r \rightarrow \infty} V^i(r, H, t) = 0$$

Figure 2 summarizes the remaining conditions, which establish the boundaries of the various circumstances affecting the termination decision.

3. At any time t , the mortgage value satisfies the following conditions:

Let $V^i(r, H, t^+) = V^i(r, H, t + 1) + MP$, then

$$V^i(r, H, t) = \begin{cases} V^i(r, H, t^+) & \text{if } H(t) > Hdn \text{ and } U(t)(t + X_i) \\ & > V_i(r, H, t^+) \text{ if continued} \\ U(t) & \text{if } H(t) > Hdn \text{ and } V^i(r, H, t^+) \\ & \geq U(t)(1 + X_i) \text{ if refinanced} \\ U(t) & \text{if } H(t) \leq Hdn \text{ if defaulted} \end{cases}$$

where $U(t)$ is the principal remaining at time t . Hdn is the boundary of default, defined as the housing price times the cost of default, or $Hdn = (V^i(r, H, t^+)/(1 + d)$. X_i is the prepayment transaction costs for individual i and d is the transaction cost of default for all individuals. This boundary condition defines the default and refinancing regions in Fig. 2. When housing prices fall so low that they are exceeded by the default cost-adjusted mortgage value, the mortgagor will exercise their put option by defaulting. The refinancing region describes a situation in which the interest rate falls to the point where the mortgage value is greater than the refinancing cost-adjusted unpaid principal. In this case, the mortgagor exercises the call option by refinancing their loan. The value of the mortgage liability takes on the value of unpaid principal $U(t)$ unadjusted by transaction costs $(1 + X_i)$, because the refinancing costs are collected by the third party who services the mortgage.

4. To improve on the previous model, we have included the effect of housing prices on the termination decision

$$V^i(r, H, t) = V^i(r, H, t^+) \text{ if } H^* > H(t) > Hdn \text{ and } V^i(r, H, t^+) \leq U(t)(1 + X_i) \text{ if restrained}$$

where LTV is the loan-to-value ratio and H^* is determined at

$$U(t) + (1 + X_i) - LTV * H(t) = V^i(r, H, t) - U(t). \quad (11)$$

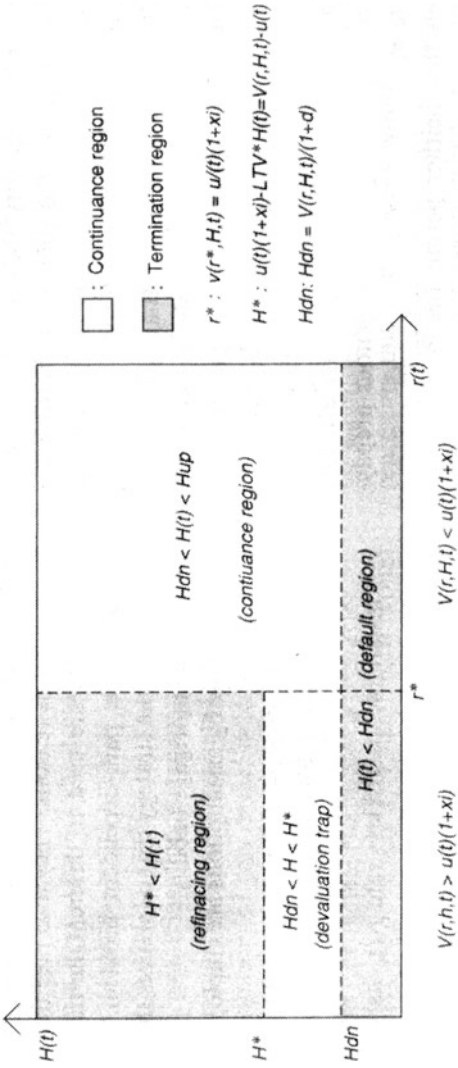


Fig. 2 Diagram of boundary conditions

This condition encompasses the devaluation trap. The devaluation trap occurs when housing prices fall between H^* and H_{dn} , where the costs of refinancing exceed its benefits. The mortgagor will be unable to refinance their loan, even though interest rates are advantageous, because they will have to pay the difference out of their pocket. And since the housing price remains above the default threshold, the mortgagor continues the mortgage. The present value of costs is determined by the left-hand side of Eq. 11, that is the difference between the unpaid principal plus refinancing transaction cost and the new loan amount, which is the housing price times the loan-to-value ratio. The benefit of refinancing is given by the right-hand side of Eq. 11, i.e., the mortgage value minus the unpaid principal. The role of the loan-to-value ratio is important in determining the size of the devaluation trap. The higher loan-to-value ratios result in decreases in the range of the devaluation trap.

Working back 1 month at a time, we can value the i th mortgage liability $V^i(r, H, t)$ by solving Eq. 10, given boundary condition 1 through 4. Given π and ρ , we can also calculate the probability that the mortgage is terminated in month t . Denote P_e the probability of termination if only exogenous prepayment occurs. Denote P_r the probability of termination if it is endogenous conditions that lead to a decision to terminate in month t . According to the survival function Eq. 5, these termination probabilities are given by

$$\begin{aligned} P_r &= 1 - e^{-(\pi+\rho)/12} \text{ if endogenous termination} \\ P_e &= 1 - e^{-\pi/12} \text{ if endogenous termination} \end{aligned}$$

We can now calculate the expected value of a single mortgage liability. That is

$$V^i(r, H, t) = \begin{cases} (1 - P_r)V^i(r, H, t^+) + P_r U(t) & \text{(if endogenous termination)} \\ (1 - P_e)V^i(r, H, t^+) + P_e U(t) & \text{(if noendogenous termination)} \end{cases}$$

50.2.2.2 Valuation of a Pool of Mortgages

To determine the value of the mortgage-backed security at any time t , as mentioned above, we can simply take the expected value of pooled mortgage liabilities

$$V(r, H, t) = \sum_{i=1}^M V^i(r, H, t) \times P(X_i = x) \quad x \in (0, 0.1] \quad (12)$$

50.3 Estimation

A model for valuing mortgage-backed securities was described that permits the determination of the security's price for given parameter values describing exogenous and endogenous factors that contribute to the termination decision. The next

logical step would be to estimate these parameter values from prepayment data. In this sector, the generalized method-of-moment technique is proposed for the estimation, where the termination probability at any given time t is required for equating the population and sample moments. In order to accomplish this, we must determine the model in terms of the probability rather than in terms of the dollar value of the security.

50.3.1 Determination of the Expected Termination Probability

In addition to equating the population and sample moments when the generalized method-of-moment technique is employed for the estimation, the calculation of termination probability is useful because it can also be utilized to determine the expected cash flows for any other mortgage-related securities, such as collateralized mortgage obligations. We first restate the procedure for determining the price in order to provide a comparison to the procedure for determining termination probability.

50.3.1.1 Procedure for Determining the Security Price

In this model, the uncertain economic environment a homeowner faces is described by two variables: the interest rate and the housing price. The term structure of the interest rate is assumed to be generated from the stochastic process described in Eq. 8 and the process of the housing prices is represented in Eq. 9. Assuming perfect capital markets, the present value $V^i(r, H, t)$ of the mortgage contract at time t is of the form

$$V^i(r, H, t) = \tilde{E}_t \left[e^{-\int_t^T r(t) dt} \bar{V}^i(T) \right], \quad (13)$$

where $\bar{V}^i(T)$ is the terminal value of the mortgage liability at expiration date T . This equation states that the value of the mortgage is equivalent to the discounted-expected-terminal payoff under the risk-neutral measure. By Girsanov's theorem, under certain circumstances, the change in measure merely produces a change in drift in the underlying diffusions. Consequently, one must substitute the risk-adjusted processes for the actual stochastic processes in Eqs. 8 and 9, which in this case are

$$dr = (\kappa\mu_r - (\kappa + \lambda)r)dt + \sigma_r d\tilde{z}_r \quad (14)$$

and

$$dH = rHdt + \sigma_H d\tilde{z}_H. \quad (15)$$

When the housing price process is transformed to its risk-adjusted form, the actual required rate of return on the house μH drops out of the equation. Therefore μH does not influence the mortgage and default option values. We know that the mortgage value $V^i(r, H, t)$ satisfies the partial differential equation specified in Eq. 10. And thus, with the appropriate terminal and boundary conditions, the value of the mortgage is determined by solving this partial differential equation (PDE) backwards in time.

50.3.1.2 Deriving the Expected Termination Probability of Mortgage i

In order to implement the parameter estimation, we are now concerned with the actual occurrence of termination instead of the dollar value of the mortgage. We begin the derivation of termination probability with the following definition:

$$\begin{aligned} P^i(r, H, t) &= \Pr((r(\tau), H(\tau), \tau) \\ &\quad \text{\textit{is in termination region of mortgage } } i, \text{ for some} \\ &\quad \tau > t, \text{ given } (r(t), H(t), t) = (r, H, r)) \end{aligned} \quad (16)$$

where (r, H, t) are the interest rate and housing price at current time t , while $P^i(r, H, t)$ is the probability that termination ever occurs beyond the current situation. The general theory of stochastic processes allows that such a probability satisfies the Kolmogorov backward equation

$$\begin{aligned} &\frac{1}{2} \sigma_r^2 r P_{rr}^i + \rho \sigma_r \sigma_H \sqrt{r} H^{\gamma/2} P_{rH}^i + \frac{1}{2} \sigma_H^2 H^\gamma P_{HH}^i \\ &+ \kappa(\mu_r - r) P_r^i + \mu_H P_H^i + P_t^i = 0 \end{aligned} \quad (17)$$

To describe the boundary and terminal conditions, we denote Ω for the part of (r, H, t) space outside the termination region, while $\partial\Omega$ forms the termination boundary. Using this notation, we have the terminal and boundary conditions

$$P^i(r, H, T) = \begin{cases} 1/M & \text{if } (r, H, t) \in \partial\Omega, t \in (o, T) \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

These conditions merely state the obvious principle that termination has a probability of $1/M$ if the conditions lead to a decision to terminate and has a probability of zero if the mortgage continues. One might recall that a pool of the mortgagors is segregated according to a discrete uniform distribution with M groups. Hence, if the environment is within the termination region, the probability of termination of any given mortgage group is $1/M$, rather than one.

The determination of probability of termination does not involve discounting, and as such the non-homogeneous rV^i term from Eq. 10 is excluded from Eq. 17. Considering we are concerned with the actual incidence of termination and not the dollar value of the termination option, the real process in Eqs. 8 and 9 are used for $r(t)$ and $H(t)$ rather than the risk-adjusted processes. Therefore, μH is required when

the probability of termination is calculated although μH has no effect on the dollar value of mortgage liability. Solving the valuation problem from Eq. 10 gives the index result of the termination region $\partial\Omega$, which consequently enters into the terminal conditions of Eq. 17. These conditions are treated as a fixed-boundary problem in the solving of Eq. 17, rather than as a free-boundary problem as in the solving of Eq. 10.

Solving Eq. 17 subject to the boundary conditions in Eq. 18 yields the termination probability at any grid (r, H, t, X_i) for the individual mortgage liability i , called P_{it} . Recalling that the hazard rate for mortgage i from Eq. 5 takes on a value of π if it is not optimal to terminate for endogenous reasons, and takes on a value of $(\pi + \rho)$ otherwise, and thus the expected termination probability for mortgage i , P_{it}^* , is calculated as

$$P_{it}^* = P_e(1 - P_{it}) + P_r P_{it} \quad (19)$$

50.3.1.3 Determination of the Expected Termination Level of Pool j

Since the distribution of the refinancing transaction cost X_i is independent of the underlying stochastic processes, the expected termination for any given pool j is calculated by counting the proportion of terminations in each transaction cost group. If we denote P_{jt}^* as the expected termination probability for a given pool j , then

$$\begin{aligned} P_{jt}^* &= \sum_{i=1}^M P_{it}^* = \sum_{i=1}^M P_e(1 - P_{it}) + P_r P_{it} \\ &= P_e \left(1 - \sum_{i=1}^M P_{it} \right) + P_r \sum_{i=1}^M P_{it} \end{aligned} \quad (20)$$

Equation 20 permits us to calculate the expected termination probability for a given pool j without having to calculate the expected probability of each individual mortgage i .

50.3.2 Estimation Approach

50.3.2.1 Generalized Method of Moments (GMM)

The generalized method of moments procedure set out by Hansen (1982) and Hansen and Singleton (1982) has been widely used in financial market applications and in labor market application. The procedure is a limited-information method analogous to two-stage least squares. GMM provides a means of estimating parameters in a model by matching theoretical moments of the data, as a function of the parameters, to their sample counterparts.

The usual way of proceeding is to identify error functions of the parameters and observable data which have an expectation of zero, conditional on the information

available at the time the data are observed. That is, if we let θ_0 denote the true vector of parameter values, there are error functions $e_{it}(\theta_0)$, $i = 1, 2, \dots, M$, satisfying the orthogonality conditions

$$E[e_{it}(\theta_0) | I_t] = 0, \quad (21)$$

where e_{it} is a function of the parameters and of data up to and including time t and I_t is the information set at time t . This equation states that these error functions have a mean zero conditional on the information set at time t , when the functions are evaluated at the true parameter value. The implication is that the errors must be uncorrelated with the variables in the information set I_t , and thus, if z_{jt} is a finite dimensional vector of random variables that are I_t measurable, then by the law of iterated expectations

$$E[e_{it}(\theta_0)z_{jt}] = 0. \quad (22)$$

If e is a $M \times T$ matrix, and z is $N \times T$, this can be rewritten as

$$E[g_t(\theta_0)] = 0, \quad (23)$$

where g is the $MN \times T$ matrix formed by taking the direct product of e and z . This equation is the basis for the GMM technique. Suppose the sample counterpart of this set of population moments is the MN -vector valued function $g_T(\theta)$, which is defined by

$$g_T(\theta)_i = \frac{1}{T} \sum_{t=1}^T g_{it}(\theta). \quad (24)$$

The usual GMM estimation procedure involves minimizing a quadratic form of the type

$$Q_T(\theta) = g_T(\theta)' W g_T(\theta), \quad (25)$$

where W is some positive-definite weighting matrix. This is usually done in two steps. First, take W to be the identity matrix and perform one minimization. Next calculate W_T , the sample estimator of

$$W_0 = (E[g_t(\theta_0)g_t(\theta_0)'])^{-1} \quad (26)$$

and use this as the weighting matrix in the second stage. As long as $W_T \rightarrow W_0$ almost surely, then the asymptotic variance matrix of the GMM estimator is

$$\sum_0 = \frac{1}{T} [(E[\partial g_t(\theta_0)' / \partial \theta]) W_0 (E[\partial g_t(\theta_0) / \partial \theta'])]^{-1} \quad (27)$$

In addition, the statistic $TQ_T(\hat{\theta})$, which is sample size times the minimized value of the objective function, $g_T(\hat{\theta})' W g_T(\hat{\theta})$, is distributed as a chi-squared random

variable with degrees of freedom equal to the dimension of $g_A(\theta_0)$ less the number of estimated parameters. This statistic provides a test of the over-identifying restrictions.

50.3.2.2 Moment Restrictions

The typical moment condition to use is the expectation of the difference between the observed prepayment level and its expected value, defined appropriately, equal zero. If we denote φ_{it} for the proportion of pool i prepaying in month t . The expected value of φ_{it} , t conditional on the information set at time t follows from above, assuming that the distribution of transaction costs among the mortgages remaining in the pool is known. If the termination probability of pool i at time t is P_{jt}^* , then

$$E[\varphi_{it}|I_t] = P_e(1 - P_{it}^*) + P_r P_{it}^*, \quad (28)$$

where $P_r = 1 - e^{-(\pi+\rho)\delta t}$, $P_e = 1 - e^{-\pi\delta t}$ are previously defined and P_{jt}^* is calculated from the previous section.

It is possible to calculate unconditional moment conditions by multiplying these conditional moment conditions in Eq. 28 and appropriate elements of the information set. Define the residual for pool i at time t as

$$e_{it} = \varphi_{it} - \bar{\varphi}_{it}. \quad (29)$$

This satisfies the following expression:

$$E[e_{it}(\theta_0)|I_t'] = 0, \quad (30)$$

where I_t' is a subset of the full information set at time t , which includes the interest rate path. Given this, it is possible to create more moment conditions as above. If z_{jt} is an element of I_t' , then $E[e_{it}z_{jt}] = 0$. However, z_{jt} may not be any variable that gives information about the actual sequence of prepayments. For example, setting z_{jt} equal to a lagged value of the prepayment level is not valid because the expected residual may be correlated with lagged prepayment levels. The implication of this is that there will be positive serial correlation in the residuals e_{it} . Hence, if the residuals are stacked in the usual way, averaging across time periods, one will have to deal with this serial correlation in calculating the appropriate standard errors for the GMM estimators.

To avoid the issue of serial correlation, the residuals can be stacked by averaging across pools, instead. Under the null hypothesis of independent pools, this way of stacking will result in no correlation between the contemporaneous residuals from different pools. Therefore, by assuming that the mortgages are drawn from a well-behaved underlying distribution, the sample estimator W_T is still a consistent estimator of the optimal weighting matrix W_0 , and the usual asymptotic standard error results are valid.

50.4 Conclusion

The valuation model of mortgage-backed securities proposed here is a model that extends the rational option pricing approach used by previous authors. This model is able to capture many important empirical regularities observed in prepayment behavior that have previously been modeled successfully using only purely empirically derived prepayment models. However, in these purely empirical models, estimation of the prepayment behavior and the valuation of a mortgage-backed security are often treated as completely separate problems. This model prevents *ad hoc* integration of the estimation of prepayment and the valuation of a mortgage-backed security, and links these two into a structured model. Therefore, this model can address economic questions that are beyond the scope of purely empirical models, while possessing a simple reduced form representation that allows estimation using observed prepayment data.

This integrated model captures the fundamental characteristics of a mortgage-backed security, such as exogenous prepayment, endogenous prepayment, transaction costs of refinancing and default, heterogeneity among mortgagors, and the issue of path dependence. In addition, the treatment of embedded options, the prepayment (call) option and the default (put) option, are modeled with care to accommodate more realistic aspects of a mortgagor's behavior. In particular, the payoff from the incident of default is modeled as an insured mortgage, so that the potential discrepancy in the pricing of the mortgage backed security is eliminated.

Another important innovation of the model is the explicit modeling of the housing devaluation effect that was the prevailing phenomenon in the early 1990s due to declined home prices. Over last several years, housing prices have been rising at unprecedented rates. A correction in the housing market is likely to occur in the near future and trigger a devaluation-induced prepayment slowdown. The term used for the devaluation effect is "devaluation trap" because the effect is activated only when housing prices fall to a degree at which the costs of refinancing exceed its benefits. The mortgagors are trapped and unable to refinance their loans even though interest rates are advantageous, because the new loans entitled from devaluated houses are no longer sufficient to cover the costs of refinancing.

Two constituents allow this model to come closer than previous models in describing empirical prepayment and price behavior. These are the incorporation of mortgagors' heterogeneity and the delaying of the rational prepayment decisions of mortgage holders. The heterogeneity of mortgagors is accomplished by introducing heterogeneous refinancing transaction costs. And the mortgagors' prepayment decisions are assumed to occur at discrete intervals rather than continuously, as was assumed with previous rational models. Hence, these two combined factors produce smoother prepayment behavior as observed in the actual data, and allow the model to generate prices that exceed par without requiring excessive transaction costs.

It is known that utilizing maximum likelihood as a means of estimating the parameters in parametric hazard models of prepayment is problematic, given the constitution of available prepayment data. Thus, by utilizing an alternative approach,

the generalized method of moments, the model parameters can be estimated. This approach overcomes the problems associated with maximum likelihood in this setting.

References

- Brennan, M., and E. Schwartz. 1985. Determinants of GNMA mortgage prices. *AREUEA Journal* 13: 209–228.
- Davidson, A.S., M.D. Herskovitz, and L.D. Van Drunen. 1988. The refinancing threshold pricing model: An economic approach to valuing MBS. *Journal of Real Estate Finance and Economics* 1: 117–130.
- Dunn, K.B., and J.J. McConnell. 1981a. Valuation of GNMA mortgage-backed securities. *Journal of Finance* 36: 599–616.
- . 1981b. A comparison of alternative models of pricing GNMA mortgage backed securities. *Journal of Finance* 36: 471–483.
- Dunn, K.B., and Spatt, C.S. 1986. The effect of refinancing costs and market imperfections on the optimal call strategy and pricing of debt contracts. Working Paper, Carnegie-Mellon University.
- Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.
- Hansen, L.P., and K.J. Singleton. 1982. Generalized instrumental variables estimators of nonlinear rational expectations models. *Econometrica* 50: 1269–1286.
- Johnson, E.T., and Van Drunen, L.D. 1988. Pricing mortgage pools with heterogeneous mortgagors: Empirical evidence. Working Paper, University of Utah.
- Kau, J.B., D.C. Keenan, W.J. Muller III, and J.F. Epperson. 1992. A generalized valuation model for fixed rate residential mortgages. *Journal of Money, Credit and Banking* 24: 279–299.
- Kau, J.B., Keenan, D.C., and Kim, T. 1993a. Transaction costs, suboptimal termination, and default probabilities. Working Paper, University of Georgia.
- Kau, J.B., Keenan, D.C., Muller, III W.J., and Epperson, J.F. 1993b. The valuation at origination of fixed rate mortgages with default and prepayment. forthcoming. *The Journal of Real Estate Finance and Economics*.
- Richard, S.F. 1991. Housing prices and prepayments for fixed-rate mortgage-backed Securities. *Journal of Fixed Income* 54–58.
- Stanton, R. 1990. Rational prepayment and the valuation of mortgage-backed securities. Working Paper, Stanford University.
- Timmis, G.C. 1985. Valuation of GNMA mortgage-backed securities with transaction costs, heterogeneous households, and endogenously generated prepayment rates. Working Paper, Carnegie-Mellon University.



The Impacts of IMF Bailouts in International Debt Crises 51

Zhaohui Zhang and Khondkar E. Karim

Abstract

The roles played by the IMF in international debt crises have long been considered controversial among both academics and policy makers. This study reviews the role of IMF bailouts in international debt crises. The literature shows that there is a statistically significant positive wealth transfer from the IMF to the international bank creditors during major event announcements. Further, the evidence indicates the existence of market informational efficiency and different pricing behavior of different groups of international bank creditors. A pertinent future research topic would be to examine whether IMF introduces the moral hazard problem into the international financial markets.

Keywords

Bank returns · Currency crisis · Equity prices · Event studies · International debt crisis · IMF bailout · LDC loans · Market efficiency

51.1 Introduction

“The roles played by the IMF in international debt crises have long been controversial among both academics and policy makers” (Zhang 2001, p. 363). Financial crises in emerging markets and their contagion effects on the global financial system over the last two decades or so – Mexico, Argentina, Brazil, and Chile in 1982–83, Brazil in 1987, Mexico in 1990–1991 and again in 1994–1995,

Z. Zhang (✉)

Long Island University, Brookville, NY, USA

e-mail: zhaohui.zhang@liu.edu

K. E. Karim

Rochester Institute of Technology, Rochester, NY, USA

Argentina in 1995, Southeast Asia and Russia from 1997 to 1998, and the Brazilian crisis along with the U.S. Congressional debate over the increase in IMF quotas during 1997–1998 – have put the IMF under an intense spotlight in the global financial environment.

The current 182-member-country IMF was founded in July 1944 “in the hope that establishing a permanent forum for cooperation on international monetary problems would help avoid the competitive devaluation, exchange restrictions, and other destructive economic policies that had contributed to the Great Depression and the outbreak of war” (Fischer 1998). The institution has evolved through the years along with a changing international financial community. Although the current role of the IMF is being challenged from both sides, by those who denounce it and those who want to expand it (Fischer 1999), the IMF’s objectives remain the same as when it was established.¹

Who needs IMF bailouts? Two sharply opposing views confront each other. Criticisms of IMF policies can be found in David Malpass (1997, 1998), Shultz et al. (1998), Schuler (1998), Sacks and Thiel (1998), and in Wall Street Journal editorial articles (Editorial Articles 1998a, b, c; 6 April, 15 April, 23 April), etc. Some even assert that the IMF caused the crises and therefore should be abolished. On the contrary, former U.S. Treasury Secretary Robert Rubin, Federal Reserve Chairman Greenspan (1998), former Treasury Secretary Summers (1998), and Rockefeller (1998), etc. argue that IMF loans are not only necessary but also the IMF needs to be strengthened.

The central discussion surrounding IMF bailouts is about the potential moral hazard problem in the international debt markets,² or put more explicitly, socializing costs versus privatizing gains. The negative views toward the IMF hold that bailout packages encourage imprudent lending behavior and that has resulted in a large amount of bad investments. There are opinions that these bad investments are largely responsible for the financial meltdowns in the troubled Asian countries. At the same time, the Western creditor banks are avoiding the negative ramifications of their bad investments in those countries by propping up their equity values with bailouts from the IMF using member countries’ (both debtors and creditors) taxpayers funds. Arguably, those international bank creditors and troubled countries’ domestic banks should bear the negative consequences caused by their imprudent lending and investments. As put in Radelet and Sachs (1998, pp. 51–52):

The mechanics of the IMF loans merit special attention, . . . the (IMF) loan packages had the direct function of providing the central bank with resources to support the payment of debts falling due, while limiting the adverse effects of such repayments on the exchange rate. In the case of Korea, the linkage between the loan package and the repayment of the foreign debts was direct and fairly automatic. . .

The supportive views toward IMF bailouts mainly emphasize the insurance against the spread of the Asian crisis to other regions, i.e. containing the contagion effects. The following quote is again from Radelet and Sachs (1998, p. 52):

The IMF has emphasized that the lending packages were intended to support stabilization, not merely to bail out foreign financial institutions. It had hoped that its role as a quasi lender of last resort would sufficiently restore market confidence that Asian governments would not need to draw down the full package of loans. If exchange rates could be stabilized and default avoided, the thinking presumably ran, private lending would revive. . .

According to this view, the IMF can deal with those troubled governments as a neutral, nonpolitical party; can contain social costs in those troubled countries as well as the danger of causing regional security problems; and can exert leverage to restructure those countries' economic systems toward a free-market system, therefore affecting and leading political systems toward more democratic ones.^{3, 4}

While it is easy to understand why some people have a certain view of IMF bailouts, the role or the existence of the IMF itself, and the conditions the IMF enforces on bailout recipients are not well differentiated in the current discussions. The negative view toward bailouts sometimes claims that the existence of the IMF bailouts fosters imprudent lending behavior that in turn contributes to the development of currency crises. However, this view does not give the IMF credit for the condition that they lay out when they make a bailout to a country in crisis. Like the Federal Reserve's discount window policy, if the benefits of obtaining the funds are not as good as they look (i.e. the banks' books must be checked as a condition for obtaining the loan), then the incentives of committing moral hazard would be greatly reduced. The condition of the bailouts is nothing else but the counterparts of the incentives of committing moral hazard.

The following argument against the international critics of the IMF with regard to the moral hazard problem is extracted from Stanley Fischer's address⁵:

To begin with, the notion that the availability of IMF programs encourages reckless behavior by those countries is far-fetched: no country would deliberately court such a crisis even if it thought international assistance would be forthcoming. The economic, financial, social, and political pain is simply too great; nor do countries show any desire to enter IMF programs unless they absolutely have to.

This point is further supported by the initial reluctance of South Korea to ask for the IMF bailout,⁶ China's accelerated reforms in its financial sector to avoid similar crisis, and other similar arguments.

51.2 Literature Review

Event studies of international debt crises are abundant in the finance literature. These studies can be categorized according to several criteria. Based on subject matter, one group of studies (Cornell and Shapiro 1986; Bruner and Simms 1987; Smirlock and Kaufold 1987) examines the impact of the "emergence" of less developed countries' (LDC) loan problems on the value of firms. Another group of studies analyzes the impact on firm values when the "solutions" to the LDC loan problems are proposed. The literature in this second group can be further divided into two subgroups

according to whether the solutions are “direct” resolutions of the crises (Demirguc-Kunt and Huizinga, 1993; Madura et al. 1993; Unal et al. 1993; Zhang 2001; Zhang and Karim 2004), or “indirect” workouts of the crises (Billingsley and Lamy 1988; Musumeci and Sinkey 1990). Based on whether events “cluster” or not, or in other words, whether event windows are overlapping due to the characteristics of the occurrence of events, the event study methodology is also different. The clustering event analysis requires the estimation of the cross-sectional correlation between firms by employing multivariate analysis, such as in Smirlock and Kaufold (1987), Zhang (2001), and Zhang and Karim (2004). The more traditional nonclustering event studies use simple portfolio aggregation approaches, such as Fama et al. (1969).

Billingsley and Lamy (1988) studied the impact of the regulation of international lending on bank stock prices with regard to the U.S. legislative events in 1983. In the wake of the Mexican moratorium in August 1982, the United States passed the International Lending Supervision Act (ILSA), and increased the U.S. quota in the IMF by \$8.5 billion in 1983. While previous studies reveal that the impact of the passage of the ILSA on bank stock prices is negative, Billingsley and Lamy (1988) carried the study further to include the impact of the introduction of the Act in the Congress on bank stock returns. More importantly, the joint impact of the passage of the Act and the increase of the IMF quota for the United States were studied and found to be positive, though the perceived benefit of a greater IMF quota is diminished by the ILSA impact. Also, they find that the risk to the banking industry is decreased as a result of the legislative events. The authors assert that the economic significance of the legislative changes to bank stockholders depends on the perceived trade-off between the benefits of increased IMF subsidization of international loan risk and the reduced opportunity to pursue such risks under the ILSA.

The related hypotheses are: one, that investors did not perceive any of the considered legislative events to include economically material information; two, that investors did not view the exposed banks differently from nonexposed banks due to the legislative changes. They find that the United States support of the IMF event produced a daily positive excess return of about 1% for the stockholders, while the cumulative effect of the introduction of the ILSA has a negative impact on the stockholders as predicted. Both events were tested on the whole sample basis.

They also find that the nonexposed banks did not react significantly to the greater U.S. support of the IMF, while the exposed banks reacted in a vigorously positive manner. The “introduction” of the ILSA had no impact on either of the two subsamples. But the “passage” of the ILSA had a significant negative impact on exposed banks while there was no impact on nonexposed banks. Also, a significantly positive relationship was found to exist between the individual BHC stockholders’ reactions and the extent of BHCs’ Latin American loan exposures for both the passage of the IMF quota increase and the passage of the ILSA.

Demirguc-kunt and Huizinga (1993) studied the impact of “direct” official credits to debt countries on returns of foreign-exposed banks. The purpose of the paper is to infer from the movement of bank stock prices the implicit transfer of official funds

(loan to the debtor countries) back to the foreign commercial banks that made the loans in the first place. Four different types of events were tested.

From October 1982 to February 1983, the IMF made loan commitments to Argentina, Mexico, Chile, and Brazil. The main result is that the stock market did not change significantly in light of the IMF loan commitments, as market investors anticipated larger commitments to the indebted neighboring countries after the commitment to Argentina was made.⁷ And even in Argentina's case, only two banks enjoyed significantly positive returns over the 3-day event period.

The hypotheses for zero coefficients are rejected for the exposed banks and for all banks together, as expected. But they also were rejected at the 10% level for nonexposed banks. This points to possible contagion. However, the hypothesis that the event parameters are equal is rejected for all three groups of banks, indicating that investors knew at least some information about each individual bank's exposure level. This, on the contrary, implies the rational pricing hypothesis. In general, the test results are not always consistent with each other and few significant results are obtained. The obscure results may be attributed to mismatching the data selection with the event periods. All three groups of banks are categorized by using the exposure data till the end of 1988. The tests could be very misleading when the actual data used were dated several years later. Also, the study left unexplained how market investors knew information about individual bank's exposures without getting it as public information.

The IMF's direct resolution of the Asian crisis in 1997 consists of clustering events that can be analyzed by conducting multivariate analyses. The IMF's bailout of South Korea in December 1997 is considered as an event that simultaneously affected all firms related cross-sectionally. The IMF's direct involvement in the global financial crisis from mid-1997 to early 1999 spurred a great deal of discussion among political leaders and economists worldwide as to whether the policies of (actions by) the IMF were appropriate to solve the crisis, and whether the existence of the institution itself was necessary at all.

Zhang (2001) examines whether the IMF bailout of South Korea in early December 1997 produced significantly positive abnormal returns in the equity values of the lending institutions. If significant positive abnormal returns occurred, then we can infer that IMF bailouts are probably generating "extra" positive wealth for the private shareholders, since potential losses without the bailouts are assumed to have negative impact on the equity values of creditors.

Zhang also examines the contagion pricing notion versus the rational pricing notion. More specifically, it is to examine whether equity prices of banks with similar foreign exposure features respond to their foreign exposure levels equally cross sectionally. If so, then equity prices change with exposure levels proportionally, which indicates the existence of rational pricing in international debt markets.

The abnormal returns of the individual banks are aggregated cross sectionally into the three portfolios based on the three bank subgroups in the study: the South Korean exposed bank group, the foreign but non-South Korean exposed bank group, and the pure domestic lending bank group. Because of the clustering of the event announcements, multivariate analysis is employed to adjust for the variance estimation to take

into consideration the cross-sectional correlation between the banks in each portfolio and between portfolios. Significantly positive abnormal returns are found for the three different bank groups on the event dates, except in one case.

The event impacts on the different bank groups are different. The South Korean exposed bank group experienced the largest positive gains among the three groups, while the foreign but non-South Korean exposed bank group did not outperform the pure domestic lending bank group. This latter result may be attributed to the fact that the lack of unison of the geographical distribution of the foreign exposure among this group of banks renders the direct comparison of their respective foreign exposures less meaningful. For banks that had no or an insignificant amount of emerging market exposure, their equity behavior may be closer to that of domestic banks than to the South Korean exposed banks.

The empirical evidence here clears the controversy regarding whether the IMF generated a wealth transfer in its bailout of South Korea in late 1997. The focus of the future discussion is not whether the IMF has generated a wealth transfer from the public funds to the private shareholders, but whether this could be avoided and how?

Zhang and Karim (2004) test the informational efficiency of financial markets related to the IMF bailout of South Korea. Informational efficiency is defined as how fast the news of the bailout announcements is incorporated into the equity pricing of U.S. banks that lend in the international debt markets. If the news of the bailout announcements were incorporated into equity prices immediately, then the abnormal returns of the foreign exposed banks would be significant on event dates but insignificant on nonevent dates. Because the IMF bailout happened in international debt markets, foreign exposed banks were directly involved, and the foreign exposure variable should be directly related to their return changes. Thus, the existence of informational efficiency can be inferred by observing whether the coefficient estimate of the foreign exposure variable is significant in terms of equity pricing on both the event and nonevent days. It has been shown in the literature that the foreign exposure variable has been the most important variable in studying equity responses in international debt crises. Presumably, if the market is informationally efficient, then the foreign exposure variable should be incorporated into the equity pricing, and its coefficient estimate is significant on the event dates, but the variable should not be incorporated into the equity pricing, and its coefficient estimate is insignificant on the nonevent dates.

For the foreign exposed banks, the mean abnormal returns tend to be significant on event dates but insignificant on nonevent dates. This evidence indicates that the news of bailout announcements was incorporated into the equity pricing immediately. There was no delay or lag effect reflected in the foreign exposed banks' equity prices. This supports the existence of market informational efficiency during the IMF bailout of South Korea in late 1997. Also, a quadratic cross-sectional regression model is employed to further examine this question by studying whether the equity returns changed proportional to the exposure levels. The CAR model on the main event date and the cumulative regression of ARs on both event dates are significant at the 5% level, while the CAR models on nonevent dates are not significant. The evidence indicates that the market is informationally efficient during the IMF bailout

of South Korea in late 1997 as investors incorporated the foreign exposure into pricing their bank equities rapidly and proportionally. There is a significant positive relationship between the banks' equity prices and their respective exposure levels on the event days. This relationship is not shown on the nonevent dates. The empirical evidence here indicates that the banks' foreign exposure information seems to be either publicly available to the markets, or investors are able to get access to this type of information.

51.3 Suggestions for Future Research

Existing literature has tried to answer the question of whether there is a potential wealth transfer from the IMF to the private shareholders of international bank creditors resulting from IMF bailouts. A more pertinent question, which is also the central debate of all the bailout events, is left unanswered. This is the issue of a potential moral hazard problem in international financial markets. The test of moral hazard requires the testing of structural risk changes before and after a bailout event. The existence of the moral hazard problem is indicated if the risk structure changes that occur after the event are significant.

While the foreign exposure variable is probably the most important variable in this type of study, the exposure data for a specific country are not available all the time. Also, due to the lack of data reporting unison, the foreign exposure examined in the literature cannot be specified according to detailed geographical locations either. The test results would certainly be improved if such data were to become available in the future.

The results would be more complete if other international lenders could be examined simultaneously along with U.S. bank creditors.

On the methodology side, while it is common to use a two-index market model in bank studies to provide the parameter estimates, which are used in the event window to calculate the abnormal returns, the significance of the contribution and the depth of the effect of the second index (usually a stationary interest rate index), in addition to the market index, needs further exploration. In other words, whether the results are sensitive to the omission of the second index or to an alternative interest rate index deserves more research effort. Also, autocorrelation is often assumed to be zero in the literature. The complexity of incorporating autocorrelation into the models so far has prevented this type of analysis being carried out in event studies. While this may not cause serious problems, the incorporation of autocorrelation into the variance estimation should give more accurate results and inferences.

Further, a random coefficient model can be employed in the estimation window to allow for any possible structural changes before the underlying event window. If multiple coefficients are obtained, then the most recent one should be used in the event window analysis. This step is especially necessary when the length of the estimation window is long.

51.4 Notes

1. See the IMF's Articles of Agreement online at: <http://www.imf.org/external/pubs/ft/aa/index.htm>.
2. Radical differences also exist as to whether the Asian countries' markets have been opening enough or whether the countries should strengthen the free-market aspects of their economies before they open further. Related discussions also focus on the transparency and regulation issues in those troubled countries (Camdessus, January 16, 1998a).
3. For example, Summers (1998) said that: "The IMF has a unique ability to provide apolitical, conditional finance... in the context of strong reforms."
4. Camdessus's (January 22, 1998b) address at Transparency International: "The IMF helps members impose the management of their public resources and establish a stable and transparent regulatory environment for private sector activity, a *sine qua non* for economic efficiency and the eradication of corruption."
5. Same as endnote 1.
6. The influential *Dong-A Ilbo* newspaper claims: "The party is over, Korea's international standing has shamefully crashed," extracted from "Out of Our Hands," Lee, 1997, *Far Eastern Economic Review*, p. 81.
7. Another explanation would be that banks were not required to publish their developing country exposures at that time. This made it much more difficult for the investors to respond in any meaningful way.

References

- Billingsley, R.S., and R.E. Lamy. 1988. The regulation of international lending: IMF support, the debt crisis, and bank stockholder wealth. *Journal of Banking and Finance* 12: 255–274.
- Bruner, R.F., and J.M. Simms. 1987. The international debt crisis and bank security returns in 1982. *Journal of Money, Credit and Banking* 19 (1): 46–55.
- Camdessus, M. 1998a. An Address at the Press Conference in Kuala Lumpur, 16 January. <http://www.web.org>
- . 1998b. The IMF and Good Governance. addressed in Paris on 21 January. <http://www.imf.org/external/np/speeches>
- Cornell, B., and A. Shapiro. 1986. The reaction of bank stock prices to the international debt crisis. *Journal of Banking and Finance* 10: 55–73.
- Demirguc-Kunt, A., and H. Huizinga. 1993. Official credits to developing countries: Implicit transfers to the banks. *Journal of Money, Credit, and Banking* 25: 76–89.
- Editorial Article. 1998a. What's an IMF for? *The Wall Street Journal*, 6 April 14.
- . 1998b. The IMF crisis. *The Wall Street Journal*, 15 April 22.
- . 1998c. IMFonomics. *The Wall Street Journal*, 23 April 18.
- Fama, E.F., E.L. Fisher, M. Jensen, and R. Roll. 1969. The adjustment of stock prices to new information. *International Economic Review* 1–21.
- Fischer, S. 1998. The Asian crisis: A view from the IMF. addressed in Washington, DC on 22 January. <http://www.imf.org/external/np/speeches>
- . 1999. On the need for an International Lender of Resort. An address to AEA/AFA joint luncheon meeting in New York, 3 January. <http://www.imf.org/external/np/speeches>

- Greenspan, A. 1998. "Remarks by Greenspan." at the Annual Convention of the American Society of Newspaper Editors, in Washington, DC, on 2 April, *WSJ Interactive Edition*.
- Lee, C. 1997. The next Domino? *Far Eastern Economic Review* 20: 14–16.
- Madura, J., A.L. Tucker, and E.R. Zarruk. 1993. Market reaction to the thrift bailout. *Journal of Banking and Finance* 17 (4): 591–608.
- Malpass, D. (1997). Break the IMF shackles. *The Wall Street Journal*, 26 September 15.
- . 1998. The road back from devaluation. *The Wall Street Journal* 14: 16.
- Musumeci, J.J., and J.F. Sinkey Jr. 1990. The international debt crisis, investor contagion, and bank security returns in 1987: The Brazilian experience. *Journal of Money, Credit, and Banking* 22: 209–220.
- Rockefeller, D. 1998. Why we need the IMF? *The Wall Street Journal*, 1 May 17.
- Radelet, S., and J. Sachs. 1998. The East Asian financial crisis: Diagnosis, remedies, prospects. *Brookings Papers on Economic Activity* 1: 1–90.
- Sacks, D., and Thiel, P. 1998. The IMF's big wealth transfer. *The Wall Street Journal*, 13 March 14.
- Schuler, K. (1998). A currency board beats the IMF Rx. *The Wall Street Journal*, 18 February 19.
- Shultz, G., Simon, W., and Wriston, W. 1998. Who needs the IMF. *The Wall Street Journal*, 3 February 15.
- Smirlock, M., and H. Kaufold. 1987. Bank foreign lending, mandatory disclosure rules, and the reaction of bank stock prices to the Mexican debt crisis. *Journal of Business* 60 (3): 77–93.
- Summers, L. (1998). Why America needs the IMF. *The Wall Street Journal*, 27 March 16.
- Unal, H., A. Demirguc-Kunt, and K. Leung. 1993. The Brady Plan, 1989 Mexican debt-reduction agreement, and bank stock returns in United States and Japan. *Journal of Money, Credit and Banking* 410–429.
- Zhang, Z. 2001. The impact of IMF term loans on U.S. Bank creditors' equity values: An event study of South Korea's case. *Journal of International Financial Markets Institutions and Money* 11: 363–394.
- Zhang, Z., and K.E. Karim. 2004. Is too-big-to-fail policy effective for US Banks in an international currency crisis? *Review of Pacific Basin Financial Markets and Policies* 7: 311–333.



Corporate Governance: Structure and Consequences

52

Bikki Jaggi

Abstract

We discuss in this chapter important aspects of corporate governance structure and examine its impact on corporate performance. We especially focus on the internal and external control mechanism and discuss how they make corporate governance more effective, enhance the quality of financial reporting, improve firm performance, and thus enhance firm value.

With passage of the Sarbanes-Oxley Act in 2002 (SOX), two important changes have taken place in the internal control mechanism. First, the audit committee has received increased attention; it has been given authority to appoint external auditors and to deal with them directly on accounting issues. Moreover, it has the responsibility to ensure high quality financial reporting. Second, internal controls have been strengthened to provide an effective monitoring of managerial activities. The main objective of internal controls is to ensure that managerial activities are properly supervised and managers do not use the flexibility provided in accounting standards to achieve their personal goals that are inconsistent with investors' goals. In order to enhance the quality of financial reporting, managers are especially monitored to ensure that they do not engage in policies and activities that result in manipulation of reported earnings.

The SOX has also improved the external control mechanism, provided by external auditors and market controls. The SOX especially focuses on ensuring independence of external auditors by restricting their functions to auditing only and not permitting them to perform advisory and other forms of nonauditing services, with the exception of tax services, so that their independence is not compromised. Additionally, the auditor in-charge is rotated under SOX to ensure

B. Jaggi (✉)

School of Business, Rutgers University, Newark-New Brunswick, NJ, USA

e-mail: jaggi@business.rutgers.edu

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_52

1187

his/her independence. We also discuss different antitakeover control devices, which are triggered when a firm becomes a target for takeover because other firms perceive the target's weak performance as an opportunity for takeover and benefit from it.

Keywords

Audit committee · Board independence · Board size · Corporate board structure · External auditor · External control mechanism · Internal control mechanism · Nomination committee · Sarbanes-Oxley act · Two-tier corporate structure · Unitary corporate structure

52.1 Introduction and Framework for This Chapter

In this chapter, we discuss important aspects of corporate governance structure and also discuss rules and regulations, especially those introduced by the Sarbanes-Oxley Act 2002, to improve the effectiveness of corporate governance. In Sect. 52.2, we explain what is meant by corporate governance and discuss its importance in public companies. Additionally, in this section, we discuss certain aspects of Sarbanes-Oxley Act 2002 which focus on improving corporate governance of US companies. Important characteristics of corporate structure are discussed in Sect. 52.3. In Sect. 52.4, we examine how corporate governance functions in the US companies and we focus on the role of different committees in making it effective. In Sect. 52.5, we explain the need for internal controls and focus on the requirements under the Sarbanes-Oxley Act 2002. The use of market control mechanism for smooth functioning of corporate governance is examined in Sect. 52.6, and the effects of corporate governance on firm performance and financial reporting are briefly discussed in Sect. 52.7.

52.2 Definition and Importance of Corporate Governance

52.2.1 What Is Meant by Corporate Governance?

Corporate governance has been defined from different perspectives in the literature as well as in practice, and these approaches potentially cover a variety of distinct economic phenomena. Shleifer and Vishny (1997) define it from the investors' perspective and describe it as a mechanism in which the capital suppliers assure themselves of getting a return on their investments in corporations.¹ The Office of Economic and Corporate Development (OECD) (1999) describes corporate

¹Refer to Shleifer and Vishny (1997) for a survey on corporate governance.

governance as a system by which business corporations are directed and controlled. This concept of corporate governance includes a corporate governance framework that specifies distribution of rights and responsibilities among different participants in the corporation and a corporate structure that includes the Board, Managers, Shareholders, and other Stakeholders. Additionally, the structure spells out the rules and procedures for making decisions on corporate affairs. Some authors, however, consider corporate governance as a tool that is used to promote corporate fairness, transparency, and accountability (e.g., Wofensohn, ex-President of World Bank). Mathiesen (2002) describes it as a mechanism that investigates how to secure and motivate management by using the incentive system, including contracts, organization designs, legislation, etc.

In order to have a better understanding of corporate governance, we first briefly define the concept of Corporation. Overall, a Corporation is described as an instrument through which capital is assembled for activities that include producing and distributing goods and services and for making investments to achieve corporation's goals and objectives. In other words, it is a mechanism that allows different parties to contribute capital, expertise, and labor, for the benefit of all. This concept of Corporation creates a limited liability for investors, and it separates the corporation from its owners, managers, and employees. The limited liability leads to a limited authority for shareholders, which primarily includes the right to elect directors. The directors have the fiduciary responsibility and the management runs the corporation to increase the shareholders' wealth and protect their interests.

Based on the above description of a corporation, we define corporate governance in a broader term as follows: The corporate governance is a mechanism for directing and controlling the corporation, and it includes a governance structure that specifies distribution of rights and responsibilities of different components of the structure, which consist of shareholders, corporate board, and corporate executives. It provides rules and procedures for making managerial decisions on corporate affairs and on the monitoring mechanism over managerial decisions and activities.

52.2.2 Agency Problem and Corporate Governance

The separation between capital providers and managers gives rise to an **agency problem** in corporations. The agency problem arises because of the conflict of interests among different groups of people working in the corporation, which include the shareholders, the managers, the employees, and the outside parties dealing with corporation, such as, suppliers, creditors, clients, etc. The conflict may be between the stockholders and management, the stockholders and bondholders, and between the majority stockholders and minority stockholders. As a result of this conflict, different groups may not be able to work cohesively to achieve the company's goals. Instead, they may focus on their group's goals at the cost of other groups. Consequently, the agency conflict is not conducive to enhancing the shareholder value in a corporation. Thus, one of the objectives of corporate governance is to minimize the

effect of agency conflict on corporate performance so that the company's goal to maximize the shareholders' value is achieved.

52.2.3 Need and Importance of Corporate Governance

Corporate frauds, bankruptcies, and other financial catastrophes in 1990s that led to destruction of shareholders' wealth in billions of US dollars, loss of jobs, criminal investigation of executives, bankruptcies, etc., generated renewed interest of politicians, regulatory agencies, and academics in corporate governance. Typical examples of large corporate failures include Enron, Worldcom, Global Crossing, Bear Stern, Tyco, Adelphia, etc. Among other things, corporate failures highlighted the weaknesses in corporate governance that enabled the managers to maximize their own wealth by ignoring public companies' overall goals of protecting shareholders' interests and maximizing company value. Additionally, the impact of corporate failures on the stock market and country's economy became quite evident and it was realized that weak corporate governance could become hindrance in smooth functioning of the stock market and it could also jeopardize country's economic growth.

Recognizing the importance of corporate governance, especially for gaining investor confidence for proper functioning of the stock market, political leadership, policy makers, and regulators decided to develop certain ground rules for smooth functioning of US companies. Their efforts culminated in the passage of the Sarbanes-Oxley Act in 2002.

52.2.4 Impact of Sarbanes-Oxley Act 2002 (SOX) on Corporate Governance

The main objective of SOX is to provide rules and regulations for proper functioning of public companies and to regulate the activities of professionals associated with these companies, which include financial intermediaries, insurance companies, brokerage houses, accounting and auditing professionals, etc. Moreover, the Act provides control mechanism to improve corporations' monitoring function and provides assurances to the investment community that managerial decisions and activities are properly supervised. This Act in fact is the most important legislation after passage of the Securities and Exchange Acts in 1933 and 1934, which for the first time provided rules and regulations to legislate corporate financial activities in the United States.

New rules and regulations contained in the Act for monitoring managerial activities and decisions are especially designed to ensure that managers do not focus on the short-term goals which might help them to enhance their own wealth by maximizing their own compensation, including bonuses. The rules shall discourage managers to engage in earnings manipulation, overstatements, enhanced risk taking, providing misleading information to investors, and an outright fraud.

Moreover, the new control mechanism shall encourage managers to provide reliable and up-to-date information on a timely basis. It also requires the firms to develop information in case the firm goes in a financial distress situation, which shall be available to investors on a timely basis. Investors do not have to wait for this information until it is too late and the impact of financial distress is already having its effect on the company's stock price.

52.3 Important Elements of Corporate Governance

52.3.1 Corporate Governance Structure

There are two types of corporate governance structures and these are known as the Unitary Corporate Board and two-tier Corporate Board.

52.3.1.1 Unitary Corporate Governance Structure

The Unitary Corporate Governance Structure, there is only one Board of Directors and this board is created to look after the shareholders' interests in the company. The company management headed by the Chief Executive Officer runs the day-to-day affairs of the company, and the Board of Directors provides advisory and monitoring functions. This type of corporate system is mostly used by companies in the common law countries and especially in the USA and the UK.

Functions of the Corporate Board in this system generally are the following: (1) to fulfill the legal requirements of the corporate charter, (2) to protect shareholders' interests, (3) to counsel and advise management, (4) to review and approve corporate plans and actions, (5) to monitor management activities, (6) to assess management performance, (7) to decide on management compensation, (8) to decide on hiring and firing of top management, and (9) to serve as a link between external auditors and management.

Members of the corporate boards in this system are elected by the stockholders in the annual meeting. The slate for election is, however, prepared by the nominating committee of the existing Board of Directors, and their decision to formulate this slate is significantly influenced by the CEO's recommendations.

52.3.1.2 Two-Tier Corporate Governance Structure

The two-tier Corporate Governance Structure, also known as the Dual corporate system, consists of two boards, which are known as the Management Board and the Supervisory Board (Vorstand and Aufsichtsrat, respectively, in German language). The Management Board is responsible for the company's day-to-day operations, whereas the Supervisory Board takes care of all capital providers' interests. In addition to the shareholders' interests, it also looks after the interests of other capital providers, which include banks, financial institutions, investment companies, other corporations, etc. In view of difference in the focus of this corporate board compared to the corporate board in the unitary system, appointment of its members need not be based on their impartiality and independence. Instead, their appointment is based

on their capability to represent the interests of a particular group of capital providers. Thus, their appointment generally reflects the company's financial and commercial relationships.

In addition to the capital providers, the German Supervisor Board also includes a representative of employees so that the employees also have a voice in the Board. The two-tier Corporate Board System is generally used by the companies in the German speaking countries, and a modified two-tier structure is used by the Chinese firms.

52.3.2 Two-Tier Corporate Structure in the Chinese Firms

The Chinese public companies also follow the two-tier Corporate Board System, but it differs from the German two-tier system. Though supervisors in the Chinese firms are appointed by the controlling shareholders, their appointment is generally done by the Chinese government in most companies because the government holds the controlling interest in these companies. It has been observed that the appointment of supervisors is generally done on their political affiliations and contributions. Supervisors in the Chinese companies can be either insiders or outsiders.

A significant difference in the two-tier system in the Chinese companies compared to the German companies is the authority of the Supervisory Board to appoint members of the Management Board. Whereas the German Supervisory Board has the authority to appoint or dismiss the members of the Management Board, the Chinese Supervisory Board has no authority to appoint or dismiss these members. The Chinese Management Board is also appointed by the Chinese government, a majority shareholder in most of the companies.

52.3.3 Independence of Corporate Boards in the Unitary System

52.3.3.1 Different Types of Directors

The corporate board directors can be classified into the following three categories: insiders (executive directors), outsiders (nonexecutive independent directors), and grey (directors with some interest in the firms, such as bankers, attorneys, etc.). An important question in this regard is how important it is to have an independent corporate board in the Unitary Corporate Structure? In other words, should a majority of these directors be independent outside directors?

52.3.3.2 Corporate Board Independence and Corporate Failures

The board independence started attracting increased attention only recently after failures of large US corporations. It is argued that an important reason for corporate failures has been the lack of corporate board independence, which resulted in inadequate monitoring of managerial decisions and activities. In most cases, the CEO conducted the company affairs and the board did not provide effective monitoring of managerial decisions and activities. In fact, the corporate board in most

cases has been quite passive and existed in form only, and the board members never questioned the managerial decisions. Apparently, the board members in most cases were appointed based on their personal relations with the CEO or based on their interest in the company in some form, which impaired their independence and took away the incentive to question the managerial decisions. Thus, the main force behind their appointment probably has been their loyalty to the appointing authority.

52.3.3.3 Positive Aspects of Corporate Board Independence

It is now commonly recognized that independence is an important factor for the board to be effective. In fact, it is argued that a majority of members in the board shall be independent of the CEO, meaning that a majority of board members shall be nonexecutive independent outside directors. Some important arguments supporting the appointment of independent directors are as follows: First, it is argued that independent outside directors are expected to be more effective in monitoring managerial decisions and activities because they will be free of managerial influence, especially the CEO. Second, these directors will be in a better position to provide advice and guidance to the management based on their experience at other firms. Their expertise and experience gained at other companies will prove to be beneficial to the firm. Third, monitoring of managerial decisions and activities by independent board members is likely to improve the quality of financial disclosures. As a result of effective monitoring, managers will refrain from manipulating the reported earnings to meet their desired goals, and consequently it will improve the quality of reported information. Fourth, independence of corporate board members shall add to the credibility of financial information disclosed by firms, especially earnings announcements. Fifth, effective advice and guidance from independent board members with extensive experience in other firms will enable managers to improve the firm's operating performance. Sixth, better operating performance is likely to result in better market performance for the firm and this will enhance firm value.

52.3.3.4 Negative Aspects of Corporate Board Independence

Several empirical studies have evaluated whether independent boards will result in better operating and market performance for the firms. These studies, however, provide mixed findings. The lack of empirical evidence supporting the effectiveness of independent corporate boards is probably because of the weaknesses associated with board independence. First, it is argued that the board's effectiveness depends on the reliability of information on which the board makes the decisions. In the absence of executive directors, there is likely to be a higher information asymmetry. The availability of information to the board is likely to be limited. In fact, there will be lack of inside information.

Second, motivation of independent directors to provide effective monitoring and advice is questionable because the outsider directors are likely to lack motivation to devote time and energy to understand the problems and issues facing the company and find solutions to these issues and problems. Moreover, there is no legal responsibility for the outside directors to provide effective monitoring and advice.

Third, their busy schedule may also add to the lack of their motivation to solve the company's problems. It is argued that in addition to taking care of their own company, outside directors serving on several other boards will have very busy schedule, which will have a negative impact on their performance as a director of other companies.

Fourth, their independence may be questionable because their appointment will depend on their relations with the firm's CEO. Thus their appointment is likely to be based on their friendship with the firm's CEO or it will be influenced by certain obligations (e.g., working on each other boards), which will weaken their independence and effectiveness.

52.3.3.5 Corporate Board Independence and Ownership Structure

Independence of outside directors is also likely to be influenced by the ownership structure of the firm. Majority ownership of the firm can either be in the hands of the controlling shareholders or it can be widely dispersed and there are no controlling shareholders. The first type of firms is known as family-owned and controlled firms, whereas the second type of firms are public companies with widely dispersed ownership. The agency problem differs between the two types of firms. In the family firm, the agency conflict is between the controlling and noncontrolling shareholders, whereas the agency problem in the widely dispersed ownership firm is between shareholders and managers.

Independent directors in the family firms are appointed by the controlling shareholders and their loyalty is likely to be with the appointing shareholders and they are not likely to focus on the interest of the minority shareholders. This will result in an ineffective monitoring even if the board members are outsider independent board. Thus, effectiveness of independent boards in the family-controlled firms is questionable. In fact, this situation is likely to provide greater opportunities to the controlling shareholders to usurp formal powers and transfer wealth from the minority shareholders to the majority shareholders. In other words, majority shareholders will have a greater opportunity to adopt opportunistic behavior and enrich themselves at the cost of minority shareholders. Thus, outside shareholders of the family-owned firms shall be careful in evaluating the effectiveness of independent corporate boards in this type of firms.

52.3.3.6 Role of Executive Directors on Corporate Boards of Public Companies

In order to complete the discussion on corporate board independence, we would like to point out the advantages of the Executive Directors on the corporate boards. Among other factors, the Executive Directors are more knowledgeable in the company affairs, they have better inside information, and they have no time constraint to devote their attention to company affairs. These advantages of Executive Directors suggest that a completely independent corporate board will not be desirable. The presence of Executive Directors on the board will improve the Board's effectiveness.

52.3.3.7 CEO Duality and Corporate Board Independence

The CEO duality also plays an important role in ensuring independence of corporate boards. The Chairman of the Corporate Board can also serve as the Chief Executive Officer (CEO). If both positions of the Chairman of the Corporate Board and CEO are held by a single individual, that is, when CEO is also the chairman of the corporate board, the situation is termed as CEO duality. An important question facing regulators, corporations, investors, and academics is whether CEO duality is good for corporations or not.

Will it an impact on the monitoring effectiveness of the board? Will it improve firm performance? Historically, the CEO duality has been a common practice in the corporate world. An entrepreneur who starts the corporation generally holds both positions in the beginning, and this situation does not create any concern for investors. When corporations become large and ownership gets widely diffused, the CEO duality is considered undesirable because it can weaken the monitoring mechanism or it may not be good for corporate performance. The positives and negatives of CEO duality are briefly discussed below.

52.3.3.8 Positive Aspects of CEO-Duality

There are several positive aspects of CEO Duality, which make it desirable under certain circumstances. First, when the positions of chairman of corporate boards and CEO are held by a single individual, there is a lower conflict and the individual is able to devote his/her full energies to the job and there are less diversions resulting from the conflict between the chairman and CEO. Consequently, we can expect an improvement in the corporate performance and also in the board effectiveness. Second, the CEO is not concerned about any antagonism from the corporate board and especially from the board chairman. Thus, this situation is likely to result in a better working environment which will be conducive to the operating and market performance of the firm. It is generally observed that the CEO duality performs a useful function in small firms and also in the firms where ownership is concentrated in a few hands.

52.3.3.9 Negative Aspects of CEO-Duality

As the corporations become large and there is no concentration of ownership, monitoring of managerial activities becomes more important. In order to provide effective monitoring over managerial decisions and activities and to ensure that management are working in the best interest of shareholders, it is important for corporate boards, especially the board chairman, to be independent of the CEO. It is argued that in the case of large corporations, CEO duality is not likely to be conducive to ensure the shareholders' interests. Instead, the CEO duality is likely to empower the CEO to make decisions which may not be in the best interest of shareholders and instead the managers will be more concerned to enhance their own interests and personal wealth. In order to ensure that management is not working against the shareholders' interest, an effective monitoring mechanism will be needed and one important aspect of this mechanism will be an independent corporate board,

headed by an independent chairman. Furthermore, the CEO duality is likely to aggravate the agency problem.

52.3.3.10 CEO Duality and Institutional Shareholders

The institutional investors, especially CALPERS (California Public Employee Pension Fund), are especially against CEO duality because they are concerned that CEO duality would impair board independence which will not be conducive to enhance the shareholders' value. Some corporations agree with CALPERS and have given the responsibilities of the board chairman and Chief Operating Officer to two separate individuals, whereas other corporations believe that CEO duality is in the best interest of the firm because it results in better performance. The latest trend is, however, toward splitting the responsibilities of the board chairman and CEO.

52.3.4 Specialization of Independent Directors

Will specialization of independent directors on corporate boards improve the board's effectiveness? Examples of specialization of independent directors are bankers, attorneys, accountants, etc. The Sarbanes-Oxley Act requires that at least one audit committee member be a financial specialist. This means that at least one independent director should be a specialist. A question arises whether independent directors with specialization improve the effectiveness of the corporate boards. It can be argued that some specializations will be beneficial in improving the board's effectiveness. For example, a banker on the board will enable the board to provide an effective advice on dealing with banks on financing and other financial transactions, and he/she will also provide an effective monitoring function. Similarly, legal specialization will be beneficial with regard to the legal affairs of the company. Agrawal and Chadha (2005) provide evidence that a financial expert in an audit committee is likely to limit the likelihood of restatements.

52.3.5 Nomination of Outside Directors: Investors' Rights

Directors in the unitary corporate governance system are nominated by the nominating committee and are elected by shareholders in annual meetings. Based on the recommendation of the nominating committee, the existing board of directors prepares a slate for election of directors and this slate is sent to shareholders in the proxy statement. The nominating committee prepares the slate based on nominations by the existing directors and the Chief Executive Officer. Nominations can also be made by shareholders, but the Board is under no obligation to include nominations from shareholders on the slate for election of directors. As a general practice, nominations from investors are ignored.

The shareholders have the right to vote for or against the nominated director. If he/she director fails to obtain a simple majority, he/she is not elected. The candidate director rejected by the shareholders can, however, be re-nominated.

The shareholders have the right to reject the candidate second time. If the rejected director is nominated third time by the board of directors, investors have no right to reject the candidate for third time.

This process of nomination has been criticized by activists because it does not give any right to investors to nominate a director unless the existing board approves it. This issue has been debated for a long time. After a careful consideration of the issue, the SEC issued a new rule in 2010 that investors with 5% shareholdings have the right to nominate a director on the board (with or without the approval of the existing board). This rule was challenged in the court by the US Chamber of Commerce and Business Roundtable; they charged that the rule was arbitrary and capricious, violated the Administrative Procedure Act and the SEC failed to properly assess the rule's effect on "efficiency, competition and capital formation" as required by law. The US Court of Appeals for the District of Columbia Circuit rule in 2011 that the SEC rule was issued without an appropriate cost-benefit analysis of the rule. The court asked the SEC to reconsider this issue and evaluate costs and benefits involved in allowing investors to nominate directors.

52.3.6 Corporate Board Size

One of the important characteristics of a corporate board is the board size. What is an optimal size of a corporate board? This issue has been extensively discussed in the literature and practice, and there is no consensus on this issue. Most of discussion on this issue has been focused on whether a small or a large board is more effective. It is generally argued that the board should be small enough to conduct the business effectively, yet it should be large enough so that different individuals can be assigned to different committees and different viewpoints can be presented in the board.

52.3.6.1 Effectiveness of Large Versus Small Corporate Boards

It is argued that a large board possesses more capabilities for problem solving and it may allow individual directors to be more effective in their committee assignments because they are not likely to be over-burdened with such assignments. Jensen (1993), on the other hand, argues that keeping the board small will help in improving the firm performance, and he is of the opinion that if a board gets beyond seven or eight members, it is less likely to function effectively and this situation would make it easier for CEOs to control the board. Lipton and Lorsh (1992) further argue behavior norms in large boards can make them dysfunctional, and moreover, large boards will lack group cohesiveness.

52.3.6.2 Corporate Board Size and Firm Performance

Hermalin and Weisbach (2003) provide empirical evidence that the board size is negatively related to both firm performance and the quality of decision making. Similarly, Yermack's (1996) findings also show a negative relationship between the firms' market valuation and board size. Eisenberg et al. (1998) document an inverse relation between the board size and profitability for small and midsize companies in

Finland. Carline et al. (2002) provide similar evidence for the UK firms. But Van Ees et al. (2003) find no evidence on the relationship between the board size and firm performance in Netherlands.

52.4 Functioning of Corporate Boards

In this section, we discuss how the US corporate boards work in actual practice, and we discuss committee structure, meeting frequency, busy boards, and gender participation.

52.4.1 The Committee Structure

In order to be effective, corporate boards appoint several committees, some of them are standing (permanent), while others are ad hoc committees. All corporate boards have generally the following standing committees: audit committee, compensation committee, and nomination committee. We discuss the important aspects of these three committees.

52.4.2 Audit Committee

Prior to SOX, it was optional for firms to establish an audit committee in the firm. The main function of the audit committee prior to SOX was to serve as a link between the CEO and external auditors. The committee primarily provided advice to the management and assisted the management in resolving the differences between auditors and management.

The nature of audit committee has drastically changed with the passage of SOX. It is now mandatory to have an audit committee and SOX has mandated certain functions and responsibilities must be assigned to this committee. The SEC has provided certain rules to improve disclosures related to this committee, which include composition of committee, committee functions, procedures for communication between audit committee and management, etc.

The NYSE and NASDAC sponsored the Blue Ribbon Committee in response to an increasing sense of urgency surrounding the need for responsible financial reporting. Based on recommendations of this committee, the US stock exchanges now also require that all listed companies must establish an audit committee.

Under the revised SEC and NYSE rules, audit committees' primary function is to oversee firm's financial reporting process and prevent fraudulent accounting statements. Thus, the audit committee plays a critical role in financial reporting by overseeing and monitoring the management's decisions and activities and by supervising independent auditors' participation in the financial reporting process. The audit committee is also entrusted with the responsibility of resolving differences between outside auditors and management with regard to application of the

Generally Accepted Accounting Principles. In order to do so, the committee is required to meet regularly with the firm's outside auditors and internal financial managers to review the corporation's financial statements, audit process, and internal accounting controls.

It is argued that new dynamics of capital markets have presented the companies with an increasingly complex set of challenges. An important challenge is that the companies are under increasing pressure to meet earnings expectations. Regulators and investors have become increasingly concerned about inappropriate "earnings management," the practice of distorting true financial performance of the company. The changes in the market and the increasing pressures on companies to maintain positive earnings trends have highlighted the importance of strong and effective audit committees. Thus, the audit committee now plays a critical role in the financial reporting system by overseeing and monitoring the management and also by supervising independent auditors' participation in the financial reporting process.

52.4.2.1 Establishment of an Audit Committee

We highlight the rules and regulations for the establishment of audit committees and their functions and composition and also discuss whether specialization of the audit committee member will improve the committee's effectiveness. We also briefly discuss the role of the audit committee in maintaining the complaint channels in the firm.

The term "audit committee" is defined as a committee (or equivalent body) established by and amongst the board of directors of an issuer for the purpose of overseeing the accounting and financial reporting processes of the issuer and audits of the financial statements of the issuer (SOX). If no such committee is established, the entire board of directors serves as the audit committee.

52.4.2.2 Functions of an Audit Committee

The audit committee is responsible for the appointment, compensation, and oversight of the work of any registered certified public accountants' firm employed by the entity. The registered public accounting firm shall report directly to the audit committee, and it will have to resolve disagreements between management and the auditor on financial issues and reporting.

The audit committee is required to preapprove all auditing services to be provided by an external auditor. These services may entail providing comfort letters in connection with securities underwritings or statutory audits required for insurance companies for the purposes of State law and also nonaudit services to be provided by the auditor of the issuer. It is also required to preapprove all nonaudit services, including tax services, if they are not specifically prohibited to be performed by an auditor of the issuer.

52.4.2.3 Waiver of Preapproval of Nonaudit Services

The preapproval requirement is waived with respect to the provision of nonaudit services for an issuer, if the aggregate amount of all such nonaudit services provided

to the firm constitutes not more than 5% of the total amount of revenues paid by the firm to its auditor during the fiscal year in which the nonaudit services are provided, or such services were not recognized by the firm at the time of the engagement to be nonaudit services. The performance of such services is promptly brought to the attention of the audit committee of the firm and approved prior to the completion of the audit by the audit committee or by one or more members of the audit committee who are members of the board of directors to whom authority to grant such approvals has been delegated by the audit committee.

The audit committee may delegate to one or more designated members of the audit committee who are independent directors of the board of directors, the authority to grant preapprovals. The decision of any member to whom authority is delegated shall be presented to the full audit committee at each of its scheduled meetings. The registered public accounting firm that performs audit for the firm is required to report to the audit committee all critical accounting policies and practices used, all alternative treatments of financial information within generally accepted accounting principles that have been discussed with management, ramifications of the use of such alternative disclosures and treatments, and the treatment preferred by the registered public accounting firm. Moreover, the public accounting firm is also required to report any other material written communications between the registered public accounting firm and the management of the issuer, such as any management letter or schedule of unadjusted differences.

52.4.2.4 Independence of an Audit Committee

The SOX requires that each member of the audit committee of the issuer shall be a member of the board of directors of the issuer and shall otherwise be independent. In order to be considered to be independent, a member of an audit committee may not, other than in his/her capacity as a member of the audit committee, the board of directors, or any other board committee in the firm, accept any consulting, advisory, or other compensatory fee from the firm. Moreover, the audit committee shall not be affiliated in any way with the firm or any of its subsidiaries.

The SEC is, however, authorized to exempt a particular relationship with respect to audit committee members, as it determines appropriate in light of the circumstances. This exception implies that all members of the audit committee may not be independent.

Under the revised listing standards of the NYSE, AMEX, and NASD, under exceptional and limited circumstances, companies may appoint to their audit committee one director who is not independent if the Board determines that membership on the committee by the individual is required by the best interests of the corporation and its shareholders, and the Board disclose, in the next annual proxy statement subsequent to such determination, the nature of the relationship and the reasons for that determination.

Small companies are, however, exempt from disclosing reasons for non-independence of a committee member. All companies are, however, required to disclose whether audit committee members are independent or not.

52.4.2.5 Financial Expertise in an Audit Committee

The SOX requires that at least one member of the committee shall be financial expert. Furthermore, it requires that the issuer should disclose whether or not, and if not, the reasons for not having at least one member who is a financial expert. The term “financial expert” is defined through education and experience as a public accountant or auditor or a principal financial officer, comptroller, or principal accounting officer of a firm. The individual should have held a position that involved performance of functions to showing an understanding of generally accepted accounting principles, an understanding of financial statements, experience in auditing the financial statements, application of accounting principles, accounting for estimates, and experience with internal accounting controls. Overall, the regulations require that the individual should have an understanding of audit committee functions.

52.4.2.6 Disclosures on Audit Committee

The SEC adopted new rules and made amendments to its current rules requiring the companies to include certain disclosures about composition of audit committees and reports from audit committees in their proxy statements. The firms are required to inform the shareholders of the audit committee’s oversight with respect to financial reporting and underscore the importance of that role in the proxy statement. This disclosure should include that (1) the audit committee has reviewed and discussed the audited financial statements with management, (2) the audit committee has discussed with the independent auditors the matters required to be discussed by SAS 61, as may be modified or supplemented, and (3) the audit committee has received written disclosures and letter from the independent auditors as required by ISB Standard No. 1, as may be modified or supplemented, and has discussed with the auditors the auditors’ independence. The companies must disclose in their proxy statements whether their audit committee is governed by a charter, and if so, include a copy of *the charter as an appendix to the proxy statement at least once every 3 years. This should help investors to understand the role and responsibilities of the audit committee.*

52.4.3 Nominating/Governance Committee

The purpose of the Corporate Governance Nominating Committee is to assist the Board in identifying qualified individuals to become Board members and nominate directors to serve on the board and board committees. Additionally, this committee also assists the Board in assessing the Board’s effectiveness and it may also make recommendations to the Board for any improvement in the company’s corporate governance.

52.4.3.1 Composition of a Nominating Committee

The committee generally consists of at least three directors. The chair and members of the committee are appointed by the Board and they serve at the Board’s discretion. All members of the committee are generally expected to be independent outside directors.

52.4.3.2 Responsibilities of a Nominating Committee

Authority and responsibilities of this committee are as follows: (1) to organize search for individuals who are qualified to become members of the board, (2) to retain a search firm to assist the committee in the search process for qualified candidates to become directors, (3) to evaluate the suitability of potential candidates based on the qualifications of the candidates and the board's needs based on the current board's composition, board's need for expertise, diversity, and balance between inside and outside independent directors, (4) to make recommendation to the Board on individual candidate for nomination, and (5) to keep itself abreast of trends and best practices in corporate governance and review the company's corporate governance guidelines periodically.

52.4.3.3 Meetings and Attendance by the Company Officers

The committee is required to meet at least two times a year, or more often, if circumstances require. The committee may invite to its meetings any director or officer of the company to assist it in performing its responsibilities.

52.4.3.4 Annual Report

The committee shall conduct and present to the Board an annual performance evaluation of the committee. The committee shall review annually the adequacy of the committee's charge and recommend any changes that it deems appropriate to the Board for approval.

52.4.4 Compensation Committee

Traditionally, companies were not obligated to have compensation committees, and it was open to the board as a whole to perform the duties of the compensation committee. As a practical matter, most boards did not have compensation committees. But lately the need for this committee is being increasingly felt and laws/regulations in some states are making it a mandatory committee. Legislation or regulation in many jurisdictions requires that a compensation committee be established and maintained and that the majority of members be independent.

52.4.4.1 Need for Expertise on Compensation Committee

While financial expertise and experience is required for an audit committee, there is no such requirement for the compensation committee. In actual practice, however, knowledge and experience in compensation and human resources issues is considered helpful for the committee to discharge its functions. Some firms also consider committee member's qualifications to recognize investor sentiment toward particular sort of compensation arrangement.

52.4.4.2 Responsibilities of a Compensation Committee

The committee's responsibilities are decided by the board and they are generally disclosed by the company so that all concerned have a better understanding of the

role of compensation committee in the organization. Generally, the Board's mandate should be substantially broader than simply dealing with establishing senior management compensation levels.

The compensation committee can be assigned the following functions: (1) to establish CEO compensation, (2) to develop compensation philosophy for the company, (3) to assist with and review the compensation discussion and analysis, (4) to oversee equity compensation grant policy, (5) to assist the board in assessing and evaluating the CEO's performance, review and recommend the CEO's compensation, including salary, incentives benefits, and other perquisites, (6) to recommend to the board the amount, determination, and payment of remuneration to directors, (7) to report executive compensation as required in public disclosures statements, (8) to retain and terminate outside experts to deal with compensation issues, and (9) to evaluate shareholder proposals related to executive compensation.

A tool for the committee in reviewing the structure of their company's incentive programs is to evaluate the risk-taking behavior of managers. The committee shall discourage programs that involve excessive risk taking.

52.4.4.3 Compensation Committee Procedures

It is important that suitable procedures are developed and implemented so that the compensation committee can function effectively and perform useful function in the organization. The corporate secretary can be helpful in formulating these procedures.

52.4.5 Frequency of Corporate Board Meetings and Board Effectiveness

The number of times a board meets varies from company to company depending on the nature of business, complexity of business, role of CEO, and general policy of companies. There is no optimal number of meetings which a board should follow. It is, however, important that the board meetings are sufficient enough to deal with important issues of the board. If the meeting frequency is low, important issues will not get the Board's attention what they deserve. In this case, the board will be making decisions without fully considering pros and cons of the issue under discussion.

The Board's effectiveness, among other factors, will also depend on an appropriate frequency of its meetings. If the Board is not meeting when decisions need to be made, the CEO will make these decisions without the Board and this will impair the Board's effectiveness. The CEO will also not be happy if proper advice and counsel is not available when needed and this will force the CEO to make his/her decisions.

In 2002, the average S&P 500 board met 7.5 times, whereas it met 8.2 times in 2001. It is, however, reported that the frequency of board's meeting is lower when the board committees are effective.

52.4.6 Busy Directors

Individual outside directors generally serve on the Boards of several different corporations. It has lately been observed that sometimes a director serves on four or five boards of different companies. If a director serves on more than four company Boards, he/she can be considered a busy director. An important question arises whether a busy director can be effective on any Corporate Board and whether he/she can perform a useful function on the Board and make contribution to improve the Board's monitoring and advisory functions that will enhance firm value.

52.4.6.1 Executive of One Company Serving as a Director of Another Company

The directors of a company are generally executives of another company. In addition to their job as an executive of their own company, they have to attend the board meetings of other companies. An important question arises, why an executive of a company will serve as a director on the corporate boards of other companies. In other words, if an executive is busy in his/her own company, will he/she have enough time to serve on the boards of other companies? If not, what will motivate him/her to be a director on another company?

It is generally believed that these executives accept the director's position either because of personal relations with the executives of the company in which they accept the position or they accept the position to create goodwill for their company. It is conventional wisdom that directors will have less time to devote to the work of the board of another company on which he/she serves, especially when he/she is a director on several corporate boards. If majority of directors on the Board are busy directors, the Board will not be effective. In this situation, CEOs of company with less effective boards are likely to be more powerful and their actions will not be challenged. This means that may be a mutual understanding among executives to serve on each other boards so that they can enhance their own power.

52.4.6.2 Limit on the Number of Directorships for a Director

The issue of busy directors raises the question whether there should a limit on the number of boards on which a director can serve, so that the board effectiveness can be enhanced? It is difficult to develop any general rule to define when a particular director can be defined as a busy director and what should be the limit on the number of boards on individuals can serve.

It is, however, investors' right to know about the board composition and personal characteristics of board members, including the number directors' positions held by them. The corporations should be encouraged to disclose information on the board composition and provide as much information on directors as possible.

52.4.7 Diversity: Female Directors

Historically most directors have been middle-aged male directors in the companies, and only in exceptional cases directors were females or belonged to an ethnic minority community. Lately, there has been awareness that diversity in Corporate Boards is desirable because it can result in several advantages for the corporation. In more recent times, gender diversity on the Board and its workforce has become a key governance issue. Some argue that it is only equitable that the gender balance on the board is addressed and redressed given that half of the population consists of females, whereas the majority of Board members in a typical Board are males.

52.4.7.1 Positives for Female Directors

In addition to the gender balance, it is argued that female directors are considered to bring more strength to the boardroom because of their different life experience, their way of thinking, their patience, their way of dealing with people, and their cool behavior in dealing with the situation. Others argue that the main benefit of women being on the boards is in-depth discussion in the board, which generally results in a more consensus situation.

52.4.7.2 Trend in Female Directors on Company Boards

The awareness of female directors on the Board has been noticed in several western industrialized countries. An early exponent of women's representation in the boardroom was Norway; there is an enforced quota of 40% female directors on boards of all publicly listed Norwegian companies since 2008. Spain has also introduced an equality law in 2007 requiring companies with 250+ employees to develop gender equality plans, and the legislation became effective in 2015, which requires the Spanish companies to ensure that 40% of board members are female. The Dutch Code of Corporate Governance (2008) also advocates that the Supervisory board shall aim for a diverse composition in terms of such factors as gender and age. Similarly, the German Corporate Governance code (2009) requires the Supervisory Board to have respect for diversity when appointing the members of the Management Board. The UK Corporate Governance Code (2010) also encourages the Board to consider the benefits of diversity, including gender. This Code urges companies to have a well-balanced Board.

Diversity in the board should, however, should not be limited in meeting the requirement. Instead, female participation should be pursued to benefit the company, its shareholders, and other stakeholders. The ultimate goal should be to provide a better forum for discussion of ideas and policies from different perspectives and to provide a better monitoring mechanism that would enhance firm performance and value.

52.5 Internal Controls and Corporate Boards

52.5.1 Importance of Internal Controls

Companies generally institute internal controls to provide effective monitoring over employees' activities, and these internal controls are developed to ensure that employees are following the company's policies, managerial decisions are consistent with the best interests of shareholders, corporate activities do not violate the country's laws, and that there is no collusion among managers and employees to defraud the firm. Larger corporations especially require numerous analyses and reports to assure that there is compliance with the company's policies, that protection is provided against human weaknesses, errors, and irregularities, and that assurance is given to independent auditors on the accuracy of internal control system so that the scope of the audit work can be minimized. In the absence of adequate and effective internal controls, there is a danger of fraud, nonconformity with company policies, misrepresentation to managers for decision making, and misrepresentation to investors by management. These negative aspects of weak internal controls may result in financial difficulties for the firm, they may have a negative impact on the firm performance, and consequently they may have a negative impact on the stock prices and firm value. In the severe cases of inadequacy and unreliability of internal controls, there may a danger of corporate failure.

52.5.2 Administrative Versus Accounting Controls

We generally distinguish between two types of internal controls and these are administrative and accounting controls. The administrative controls consist of the firms' plan of organization, procedures, and tests concerned with the decision-making processes. The administrative controls become the starting point for accounting controls, which deal with safeguarding of assets, reliability of financial records, conformity with accounting standards, and reasonable assurances on accuracy and reliability of reported information. It is important that accounting internal controls are effective, and the following two conditions are the basic requirements to make these controls effective: first, an effective management leadership, good organization structure, and an appropriate budgetary process are established and second, sound accounting practices are developed in the firm.

52.5.3 Historical Perspective on Internal Controls

Historically, companies have not been required to certify whether internal controls exist in the firm and whether they are effective. Given the importance of internal controls for the reliability of information to investors and danger of bankruptcy in case of inadequate internal controls that may have a negative impact on smooth functioning of the financial markets and ultimately economy of the country, policy

makers and regulators should not leave it to managers to ensure the development and installation of proper internal controls. Therefore, from time to time, the policy makers and regulators had to intervene and develop rules and regulations to ensure the development of effective internal controls in the companies.

In 1977, the Foreign Corrupt Practices Act was enacted, which required the management to provide a report in the financial statements on the existence and reliability of internal controls in the firm. Despite certifications from managers, several firms suffered from weaknesses in internal controls that resulted in fraudulent reporting and that led to bankruptcies for some firms. Later the Treadway Commission was established and it issued a report on the factors that led to fraudulent financial reporting and also made recommendations to deter firms to make fraudulent reporting. These recommendations especially included strengthening of internal controls to reduce the occurrence of such reporting. Following the Treadway Commission Report, the Committee of Sponsoring Organizations (COSO) was formed to develop an integrated framework of internal control.

In 1992, the COSO released guidance for designing and implementing effective internal controls and argued that there would be potential benefits of having effective internal controls, which would include effective and efficient operations, enhanced reliability of financial reporting, and compliance with laws and regulations. No mandatory reporting of internal control, weaknesses except when there was a change in auditors, was imposed.

52.5.4 Internal Controls Under Sarbanes-Oxley Act (SOX), 2002

Despite the Foreign Corrupt Practices Act and COSO regulations, weaknesses in the internal control led to several corporate failures in 1990s, which included Enron, Worldcom, Tyco, etc. The 2002 Sarbanes Oxley Act (SOX) once again emphasized the importance of internal control systems and it required all companies to institute internal controls and provide management assessment on the effectiveness of these controls. In addition, the SOX required the large accelerated filers (firms with 750 million US dollars and above capitalization) and the accelerated filers (firms between 75 and 750 million US dollars capitalization) to provide an attestation from external independent auditors on the managerial assessment of internal controls' effectiveness. This Act has two important sections on internal controls and these are sections 302 and 404. We discuss the important aspects of these sections to highlight the requirements on internal controls under the existing laws.

52.5.4.1 Section 302 of SOX on Internal Controls

Section 302 describes the corporate responsibility for financial reports and requires among other things a quarterly self-assertion by the CEO/CFO, under personal liability, disclosing significant material financial information to investors and other stakeholders. The assertion should be included in the company's 10Q and 10 K, but no supporting documents or certification of independent examination is required under this section. The important aspects of this section are that the principle

executive officer or the principal financial officer must certify that he/she has reviewed the report. Additionally, their statement should include that based on the knowledge of signing officer, the report does not contain any untrue statement or there is no omission of material facts and that the financial statements and other financial information included in the report fairly present in all material respects the financial condition and the results of the operations of the issuing company. The signing officer is also required to state that he/she is responsible for establishing and maintaining internal controls and he/she has designed such internal controls and has evaluated the effectiveness of these controls. Furthermore, he/she should report the conclusion about the effectiveness of internal controls and all significant deficiencies in the design or operation of internal controls.

52.5.4.2 Section 404 of SOX on Internal Controls

The section 404 deals with Enhanced Financial Disclosures and it consists of two subsections, that is, 404(a) and 404(b). The subsection 404(a) requires that each company's annual report should include an internal control report containing the management's assessment of the effectiveness of ICFR, whereas the subsection 404 (b) requires the companies to have the auditor evaluate the effectiveness of the internal control system.

There are three important reporting requirements under this section. First, the report should state that the management of issuing company is responsible for establishing and maintaining adequate internal controls and procedures for financial reporting. Second, the report should contain management's assessment as of the end of most recent fiscal year on the effectiveness of company's internal controls and procedures for financial reporting. Third, it should contain a statement that the company's independent auditor has attested to and reported on the management's evaluation of the internal controls and procedures for financial reporting.

52.5.4.3 Applicability of Sections 302 and 404 of SOX

Both subsections 404a and 404b became effective for accelerated and large accelerated filers (\$75 million and above) in November 2004. Subsection 404(b) became effective in December 2007 for non-accelerated filers (less than \$75 million), and they are exempted from subsection 404(b) under the Dodd-Frank Act 2010. Reporting of management self-assessment under subsection 404(a) is considered sufficient for smaller companies.

52.5.4.4 Criticism of Section 404 (b)

The main criticism of subsection 404(b), especially for nonaccelerated filers, is based on the costs involved in meeting the requirements. Two surveys have been conducted to evaluate such costs. One survey is conducted by the Financial Executive Institute (FEI) and the other by the CPA firms. The FEI survey results report that audit fees increased 39% for the first year after SOX and average total cost was

reported to be \$4.36 million for the largest US companies (over five billion in revenues), and the details of the compliance costs by components are as follows: internal cost = 1.34 million, external cost = 1.72 million, and auditor fee = 1.30 million. Overall, the survey results show that all respondents believe that costs associated with compliance of section 404 exceeded the benefits.

On the other hand, the survey results conducted by the accounting firms D&T, E&Y and PWCopers reported that the implementation cost for section 404 on average per company was estimated to be approximately \$7.3 million, which represented 1/10th of 1% company revenues. On an average, audit fees were estimated to be approximately one-quarter of the total implementation costs per company, which means 1/40th of 1% of the company's revenues. The survey results further reported that the year-two implementation costs were expected to decrease approximately 46% compared to year-one costs, suggesting substantial nonrecurring start up and "learning curve" costs. Approximately \$1.9 million or 26% of the total cost is comprised of section 404 audit-related fees. It represents about 0.09% of the average revenue of the total sample companies. On average, each company identified 348 deficiencies and remediated 271 deficiencies and was expected to remediate an additional 77 deficiencies.

52.5.4.5 Response to Criticism of Section 404 (b) of SOX

California Public Employees' Retirement System (CalPERS) expressed the views that the costs associated with SOX are paid by shareholders and not by management. It was therefore the shareholders' choice if they support this cost when internal controls delivered better governance and management. It has been further argued that these costs are not significant compared to the losses suffered by investors. 55% of respondents believe that section 404 gives investors and other external users more confidence in a company's financial reports (83% respondents of large companies). In general, companies applaud the added focus on internal controls, but many respondents believe that the level of details required is impractical and bureaucratic. They argued that internal controls should be supported, but execution to the level of details required by the law was much more than necessary.

52.5.4.6 Auditing Standards and Internal Controls

The PACOB implemented Sarbanes Oxley Act Section 404 on the internal control system by issuing Auditing Statement No. 2 on October 7, 2003. In April 2006, they issued another auditing standard No. 4, which described the steps auditors should take when a company voluntarily engages them to report on the material weaknesses in the internal control system. The main objective of this standard is to obtain reasonable assurances whether the reported material weaknesses still exist. Thus, auditor's work is focused on whether the controls specified by management were designed appropriately and were operating effectively.

52.6 External Controls and Corporate Governance

52.6.1 Two Types of External Controls

External controls over the managerial decision-making process and activities are provided by independent auditors and also by the market control mechanism. We highlight a few important aspects of external controls by independent auditors and especially discuss antitakeover devices developed by the management.

52.6.1.1 External Independent Auditors

Under agency theory, external independent auditors play an important role to ensure that information reported by managers was accurate and reliable. Moreover, independent auditors also required to ensure that all assets in the firm are properly accounted for and there is no misappropriation of funds by the managers to maximize their own wealth. The Securities and Exchange Acts of 1933 and 1934 require that all companies registered with the SEC and publicly traded on the stock exchanges should have an independent auditor to attest to the accuracy and reliability of the financial information contained in the financial statements.

52.6.1.2 External Auditor's Functions

The SOX has clearly defined functions and responsibilities of independent auditors. The main function of an auditor is to provide attestation of the financial information disclosed in the financial statements. The main objective of this function is to provide credibility to the reported information. Other functions include reviewing client's accounting books, accounting policies, and accounting principles to ensure that the accounting policies and principles are applied consistently, there are no inaccuracies and/or misrepresentations, and that the financial reports are comprehensive.

After reviewing the client's accounting books, auditors form their opinion with regard to the accuracy and comprehensiveness and also investigate whether firms' financial difficulties may create problems in the future. After reviewing clients' accounting books and investigating different issues that they may detect during their audit-work, auditors may issue one of the following opinions: clean (unqualified) opinion, qualified opinion, going concern opinion, adverse opinion, and no opinion.

Hiring of Auditors and Communication with the Firm

Under new regulations, the audit committee not the management hires external independent auditors. External auditors also report back to the audit committee; they discuss with audit committee the findings of their audit work and also their differences with managers on the accounting policies and procedures.

Independence of Auditors

Independence of auditors is considered important under SOX. In order to ensure independence, the SOX regulations prohibit external auditors to perform any

nonaudit services, except for tax services. Moreover, the regulations require rotation of the partner-in-charge of audit so that the cozy relationship between auditor and client firm is avoided. Overall, SOX emphasizes two important issues with regard to the impairment of independence, and they are the nonaudit services and auditor's longer stay on the audit job with the firm.

52.6.2 Market Control Mechanism

The corporate control market, which relates to the takeover threats of the firms that are not creating value for investors, plays an important role in monitoring managerial behavior and this has a significant impact on corporate governance. Firms' weak performance makes them the targets for takeovers by corporate raiders as well as by other successful firms. These takeover threats serve as the market control over managerial and firm performance. Thus, market control mechanism serves as a function in keeping the managers focused on firms' overall goal of maximizing firm value for investors.

52.6.2.1 Antitakeover Devices

Being aware of take-over threats, managers, especially entrenched managers, take necessary steps to avoid takeovers. Thus, managers along with the Board of Directors, especially when the Boards are independent, work toward building defenses to protect their companies from the threat of hostile takeovers. In order to do so, they develop antitakeover devices, which are also known as shark repellents in the market. These devices, however, may sometimes render management and the Board of Directors less accountable. We first discuss important antitakeover defenses, and then evaluate the role of independent boards in the development of such devices.

52.6.2.2 Green Mail

This refers to a transaction between a large shareholder and a company in which the shareholder agrees to sell his stock back to the company, usually at a premium, in exchange for a promise not to seek control of the company for a specified period. This device is used when someone buys a large stake in the company and begins to make his presence known, perhaps by making noises about trying to take over the company. The management offers to buy him out at a substantial bonus over the stock's market price. This way the raiders achieve huge profits without even having to make a bid for the company, managers are able to keep their jobs, and the main losers are the shareholders. The passage of this device reflects the Board's neglect of shareholders' interests. The shareholders should be able to decide whether purchase of shares should be allowed. **Anti-greenmail** measures prevent such arrangements unless the same repurchase offer is made to all shareholders or approved by a shareholder vote.

52.6.2.3 Blank Check

It refers to the authority given to the Board of Directors to issue preferred stocks. The board of directors determines the preferred stock's voting rights, dividend, conversion, and other rights. While it can be usefully employed to enable a company to meet the changing financial needs, its most important use is, however, to implement poison pills to prevent takeover by placing this stock with friendly investors. This technique is thought to discourage accumulation of large blocks of stocks, but the net effect on the shareholder wealth is unclear.

Supermajority

It limits the shareholders' ability to amend the governing documents of a corporation. This clause in firms' bylaws will require that a supermajority vote of shareholders is needed to change the charter, make amendments to the bylaws, etc. This will eliminate managers' ability to amend the bylaws without shareholders' approval with a super majority mentioned in the firm's constitution/bylaws.

52.6.2.4 A Classified or Staggered Board

Directors are placed into different classes and they serve overlapping terms. Since only a part of the board can be replaced each year, an outsider who gains control of a corporation may have to wait a few years before being able to gain control of the board. This slow replacement makes a classified board a crucial component of the delay group of provisions and one of the few provisions that clearly retains some deterrent value in modern takeover battles (Daines and Klausner 2001). Most states, however, do not allow staggered boards after 2009.

52.6.2.5 Directors' Duties

This provision allows directors to consider constituencies other than shareholders when considering a merger. These constituencies may include, for example, employees, host communities, or suppliers. This provision provides board of directors with a legal basis for rejecting a takeover that would have been beneficial to shareholders.

52.6.2.6 Golden Parachutes

These are the severance agreements that provide cash and noncash compensation to senior executives upon an event, such as termination, demotion, or resignation following a change in control. They do not require shareholders' approval. Such payments are generally intended to deter takeovers by increasing the take-over costs. But these parachutes also ease the passage of mergers through contractual compensation to managers of the target company (Lambert and Larcker 1985).

52.6.2.7 Silver Parachutes

They are similar to Golden Parachutes in that they provide severance payments upon a change in corporate control, but they differ in that a large number of a firm's employees are eligible for these benefits.

52.6.2.8 Poison Pill

It provides the holders of this instrument with special rights in the case of triggering event, such as a hostile take-over bid. If a deal is approved by the Board of Directors, the poison pill can be revoked. If the deal is not approved and the bidder proceeds, the pill is triggered. Typically, poison pills give holders of the target's stock other than the bidder the right to purchase stock at a steep discount. This makes the target unattractive or dilutes the acquirer's voting powers.

Bebchuk (2003) argues that bylaws should force a change in the takeover defenses, especially the poison-pill bylaw provisions. He further argues that this provision denies the shareholders the right to make their own decisions. He proposes that the poison-pill provisions should be approved by a two-third majority of directors and they should expire within 3 years.

52.6.3 Antitakeover Devices and Independent Outside Directors

Some researchers and practitioners argue that because of the reputational concerns and fear of lawsuits, outside directors are likely to represent shareholders' interests effectively (Bhagat et al. 1987). But, on the other hand, it is argued that outside directors are more likely to align themselves with top management rather than shareholders because the top management has great influence over those who sit on the board. Brickley et al. (1994) empirically tested how shareholders would react to poison pill when the proportion of independent directors on the corporate boards is high or low.

They argued that if outside directors are independent and represent the shareholders' interest, the likelihood of their approving a poison pill that will harm the shareholders should decrease with a higher fraction of outsiders on the board. In this case, investors should react positively to the poison pill because the poison pill is not expected to hurt their interests. In contrast, if outside directors represent managerial interests, the likelihood of their approving the poison pill that will harm the shareholders will not vary with the fraction of outsiders on the board.

Thus, in the latter case, stock market's reaction to a poison pill will not depend on the board composition. The results of the study indicated a significant positive relation between the stock market reaction to the adoption of poison pills and the fraction of outside directors on corporate board. Thus, the results were interpreted to suggest that outsider directors do represent shareholders' interests.

52.6.3.1 Poison Pill and Firm Performance

Another issue related to poison pill examined in the literature is whether poison pills affect firms' operating performance. A widely held view is that poison pill negatively affects firm performance. Danielson and Karpoff (2006) empirically tested whether poison pills would improve firms' operating performance. Their findings suggested that firms experienced modest operating performance improvements during the 5-year period after pill adoption. These improvements occurred for a wide range of firms and were unrelated to specific adoption years or whether firms invested

heavily in R&D. Their findings are, therefore, at odd with the widely held view that poison pills negatively affect firm performance.

52.6.4 Role of Institutional Investors in Corporate Governance

The main institutional investors are pension funds, insurance companies, and investment companies. The institutional investors are especially concerned with election of directors and they believe that they should have a right to nominate directors because they hold significant shareholdings in the company. If they have no say in nominating directors for election, they might express their concern about corporate governance by selling their shares, which may not be good for the firms' stock price.

Institutional investors, however, perform important function in monitoring managerial decisions and activities because of their large shareholdings. They can also play a positive role in mitigating the agency problem. Their substantial influence on the company's management can be used to align management interests with those of shareholders. Thus, they can reduce the agency problem by possessing resources and expertise to monitor the managerial activities and providing an oversight function. Stapledon (1996, p. 17) writes that monitoring by institutional shareholders fits within a broad tapestry of devices and market forces which operate to reduce the divergence between the interests of managers and shareholders. Similarly, Diamond (1984) believes that institutional investors can solve the agency problem because of their ability to take advantage of the economies of scale and diversification and by demanding greater accountability from management. Chung et al. (2002), on the other hand, argue that there will be less opportunistic earnings management in firms with more institutional investor ownership because institutions will either put pressure on firms to adopt better accounting policies, otherwise they would unravel earnings management.

It is also argued in the literature that institutional investors' intervention can produce higher financial returns. Institutional activism can bring financial rewards, as more efficient monitoring of company management will encourage them to align their interests with the shareholders' interests, which will result in maximizing firm value and thus shareholders' wealth. Agrawal and Knoeber (1996) also emphasize that institutional investors' involvement can have a positive effect on corporate financial performance. Bethel et al. (1998) also report that active block-holders lead to enhanced shareholder value.

Some authors have, however, expressed the view that institutional shareholding may result in an increase in information asymmetry. It is argued that institutions can have access to privileged information, which will create an asymmetry between themselves and smaller shareholders. Institutions are actually not the shareholders, instead they represent the principals who are the shareholders. This situation thus creates an added agency problem. Not only does the shareholder (individual pension fund member) have to worry about the possible divergent objective of investee company management but they also have to worry about the activities of the pension fund managers. In fact, some authors argue that institutional shareholders support

management instead of shareholders. In fact, Davis and Kim (2007) report that although mutual funds are no more likely to vote with management of client versus nonclient firms, there is a positive relation between business ties and the propensity of mutual funds to vote in favor of management proposals. Pound (1988) argues that institutional investors have a tendency to help entrenched management by voting with the management team. Consequently, institutional shareholding may result in negative impact on the firm's long-term performance.

Institutional shareholdings may also lead to short-term profit maximization in order to make their returns look as healthy as possible in the short run. They may therefore pressure companies to focus on short-term profits rather than long-term profits. This can be detrimental to long-term company survival, as companies need to invest in long-term projects in order to ensure they grow and prosper in the long run.

52.7 Impact of Corporate Governance on Firm Performance and Disclosures

52.7.1 Firm Performance and Corporate Governance

The main objective of corporate governance is to provide advice to management and monitor managerial activities to improve firm performance so that shareholders' interests are protected and firm value is enhanced. Therefore, a question of interest is what type of corporate governance can help firms to achieve these goals. It is generally argued that an independent corporate will be more effective in enhancing firm value.

52.7.2 Managers-Shareholders' Conflict

Though firm performance is influenced by several factors, but the most important factor is managers-shareholders' conflict, which gives rise to the following three problems that will have an impact on firm performance:

52.7.2.1 Free Cash Flow Problem

Shareholders would prefer distribution of free cash flow in the form of dividends, but managers, on the other hand, would prefer investing cash flows instead of paying dividends, even if the investment projects are not profitable. Managers' preference is motivated by their power, prestige, and higher salaries, which are associated with managing a larger firm.

52.7.2.2 Risk Problem

The shareholders can diversify their risk through their portfolio choice, but managers cannot do so because they have a significant human capital investment in the firm. Thus, the shareholders are willing to undertake risky projects, but managers are

generally risk-averse compared to shareholders. Thus, managers are cautious and even unwilling to undertake risky projects that may latter become profitable.

52.7.2.3 The Horizon Problem

Managers, on average, are more myopic than shareholders in making their decisions, because of the pressure to produce immediate results. The “short-termism” may drive managers away from necessary maintenance and R&D expenditures and high-technology investment that have a longer payback period.

52.7.2.4 Consequences of Managers-Shareholders’ Conflict

Because of the above conflicts, some authors argue that managers may not invest in profitable projects which have long horizons. Consequently, the conflict is likely to result in lowering the operating performance of firms in the long run. To overcome these problems, corrective action is needed to focus managers’ attention on the main goal, that is, to maximize shareholders’ value through better performance. Broadly speaking, two types of mechanism are used to improve firm performance and thus enhance firm value. These are: (1) monitoring of managers decisions and activities through corporate boards, including proper advice and guidance to enhance firm performance by corporate boards, and (2) development of proper incentives by corporate boards to align managers’ interests with shareholders’ interests.

52.7.3 Monitoring by Corporate Boards and Firm Performance

It is generally argued that proper monitoring of managerial activities is expected to improve firm performance, and some authors have further argued that outside independent directors would provide more effective monitoring of managerial activities to achieve the objective of enhancing firm value.

52.7.3.1 Independent Corporate Boards and Firm Performance

It is argued that outside directors, who are supposed to take care of shareholders’ interest, can discipline management better and thus provide a more effective monitoring. An effective monitoring is expected to result in a higher level of sales, fewer employees’ turnover, higher labor productivity, lower selling, and general administrative expenses, etc. Cumulatively, these factors are expected to result in better firm performance.

Empirical evidence is, however, mixed on the role of independent directors in improving firm performance. Though some findings suggest that independent boards are positively associated with firm performance, others indicate that there is no significant association or negative association. So far as the positive aspects are concerned, it is argued that the appointment of independent directors would send positive signals to the market on firm performance. Rosenstein and Wyatt (1990), who evaluated market reaction to the announcements of outside board appointments, report that he finds significant positive excess returns around the days of this announcement. They interpret their findings to show that the announcement on

appointing an outside director on the corporate boards is associated with an increase in the shareholders' wealth. Fosberg (1989), on the other hand, reports no relationship between the outside directors and various variables used to gauge firm performance. In fact, to the contrary, he finds that firms with nonmajority of independent directors had a mean ROE 1.1% greater than the firms with majority of independent directors. Overall, the findings show that the presence of outside directors does not enhance firm performance.

Fama and Jensen (1983), on the other hand, argue that executive directors are more likely to have a positive impact on firm performance because inside executive directors are an important source of firm-specific information, and their inclusion on the board can lead to a more effective decision-making process. Morck et al. (1988), however, maintain that when there is a high level of inside ownership, the executive directors may use their voting power to maintain their jobs (entrenchment) and maximize their benefits by designing their compensation packages. This will not enhance firm performance.

Findings by Hermalin and Weisbach (1991) also show no significant association between firm performance and inside directors. They argue that inside and outside directors have their respective advantages and disadvantages. If each board is optimally weighted between insiders and outsiders, board composition will have no effect on firm performance. Rosenstein and Wyatt (1990) detect a different market reaction to different levels of insider managerial ownership. Reaction is significantly negative when inside directors' ownership is less than 5%. The reaction is significantly positive when the inside ownership level is between 5% and 25%. The reaction is insignificant when ownership percentage exceeds 25%. They interpret their findings to suggest that at low level of inside ownership, the market infers that an addition of another inside director is likely to be perceived as an attempt to entrench the existing management, that is, the existing CEO and hand-picked successor. Thus, there is a negative reaction at the moderate level of inside ownership, where the managerial interests are closely aligned with those of the outside shareholders, the expected benefits of the specialized knowledge provided by an inside manager outweigh the expected costs of increase managerial entrenchment. Thus, as a positive reaction at high levels of inside ownership, where insiders may have the voting power to insulate themselves from the scrutiny of outsiders, costs of the appointing insiders offset the benefits.

52.7.3.2 Financial Expertise on the Corporate Board and Firm Performance

The SOX requires that one of the outside directors on the audit committee should have accounting or financial expertise, but there is no direct requirement of expertise of outside directors on the corporate boards. The audit committee requirement thus implies that at least one outside director on the board should have financial expertise. Will financial expertise of an independent director lead to better firm performance? It is generally argued that the effectiveness of an independent director with expertise will depend on the nature of his/her expertise. It is, however, generally expected that independent directors with expertise are likely to make the boards more effective.

52.7.3.3 Alignment of Managers' Interests with Shareholders' Interests and Firm Performance

Corporate boards are also expected to motivate managers to enhance firm performance by aligning their interests with those of shareholders. The common technique used for this purpose is providing ownership of firm stocks to the managers. Moreover, the alignment of interests between shareholders and managers is also possible through the compensation system. The use of stock options in the compensation structure is expected to align the managers' interests with those of shareholders.

52.7.4 Financial Disclosures and Corporate Governance

Managers try to achieve different objectives through disclosures, especially voluntary disclosures, which may include the following: (1) send positive signals to the market to create a positive image for the company and (2) confirm or rebut non-official information floating in the market to counteract rumors in the market. Meaningful, reliable, and timely disclosures must meet the following criteria: accuracy, reliability, appropriateness, completeness, clarity, and timeliness. The reliability of disclosures is important to provide credibility to information so that investors can trust for their decisions.

52.7.4.1 Transparency in Disclosures

Bushman et al. (2004) emphasize that transparency is an important aspect of disclosures. They argue that users must be able to trust the reported information, which is the result of input from a multifaceted system whose components collectively produce, gather, validate, and disseminate information. Authors further differentiate between three aspects of transparency and these are: financial transparency, governance transparency, and risk transparency.

The financial transparency refers to the comprehensiveness (intensity) and timeliness of financial information disclosed. Comprehensive disclosures should meet the following three criteria:

- First, *obfuscation* should be avoided. It is argued that disclosures can result in *obfuscation* when managers possess more private information and they are less forthcoming with their overall disclosures (including voluntary disclosures). It will create higher information asymmetry. If disclosures are not comprehensive, it is feared that analysts may be more inclined to obtain information on these firms from private sources rather using the publicly available information because potential rewards from the use of private information would be higher (Barth et al. 2001). Consequently, publicly disclosed information by firms will have lower reliability.
- Second, disclosures should not be *opaque*. Many firms are guilty of providing *opaque* (hard to understand and explain) disclosures, and such disclosures are not suitable for investment decisions, especially, when the information relates to

special purpose entities, stock option expensing, intangible assets, and research and development.

- Third, disclosures should be *timely*. The main issue involved in the timeliness of disclosure relates to disclosure of losses. Should they be disclosed without any time delay or with delay? Similarly, when gains should be disclosed? The timeliness of disclosures of gains and losses is referred to as *accounting conservatism*. Conservative accounting generally provides more timely and reliable information. The reliability and timeliness of disclosures in turn determine the disclosure quality, and the disclosure quality builds credibility of disclosures in the financial markets.

52.7.4.2 Factors Influencing Transparency

Transparency of disclosures is influenced by the following factors: (1) monitoring effectiveness of the corporate boards, (2) ownership structure, and (3) political economy.

An effective corporate board should disclose more comprehensive and timely information disclosures, which should be useful to investors in making informed judgments. An independent board, with experienced and expert directors, but not very busy directors, will be considered an effective board. The board's effective will further enhance when the number of meetings it holds is reasonable so that different issues be discussed. Empirical evidence shows that independent boards are associated with higher voluntary disclosures. Chen and Jaggi (2000) show that there is a positive association between a higher percentage of independent directors and financial disclosures in Hong Kong firms.

The corporate ownership structure also plays an important role in disclosure transparency. There is no strong incentive for management to achieve high disclosure transparency when ownership is concentrated in the hands of family members, who also control firm's management. Transparency in disclosures is, however, important in firms with diffused ownership because public disclosures by these firms are expected to be more reliable and cost-effective.

The corporate disclosures are also influenced by the political economy, which represents a range of institutional arrangements that capture important relations between the government and the economy. These relationships show whether there is concentration of political power (autocracy), what is the extent of the state corporate ownership, what are the costs of entry imposed on the start-up firms, and what is the risk of expropriation.

52.7.5 Corporate Governance, Risk Assessment, and Risk Disclosures

Risk assessment and risk disclosures have recently started receiving increased attention of managers, corporate boards, and also auditors. Without an effective risk assessment, companies are unable to prioritize and efficiently deploy their resources and develop reliable monitoring activities. Risk assessment includes both an evaluation of the inherent risks (e.g., impact and likelihood analysis) and

an evaluation of the effectiveness of management's controls and risk management activities.

The goal of an effective risk assessment process is to identify coverage gaps and develop plans to remediate coverage deficiencies before they are exposed. The corporate board has multidimensional responsibilities in steering top management towards the right risk choices. The board should monitor the risk situation of the company systematically to identify and evaluate multiple sources of risk, and the following guidelines will be useful in the process of risk evaluation: (1) take a portfolio view of corporate risks, (2) be apprised more specifically of the major risks (for major risk combinations) that could significantly alter business perspectives, and (3) evaluate the way in which management has embedded risk management within the corporation, ask organizational questions, such as "do we need a chief risk officer?" "how risk is being evaluated?"

Corporate disclosures should be in compliance with legal regulations and administrative rules, and additionally they should disclose risk information voluntarily if is not covered by the regulation but is important for investors to make their decisions.

References

- Agrawal, A., and S. Chadha. 2005. Corporate governance and accounting scandals. *Journal of Law and Economics* 48: 371–406.
- Agrawal, A., and C.R. Knoeber. 1996. Firm performance and mechanisms to control agency problems between managers and shareholders. *Journal of Financial and Quantitative Analysis* 31: 377–397.
- Barth, M.E., R. Kasznik, and M.F. McNichols. 2001. Analyst coverage and intangible assets. *Journal of Accounting Research* 39: 1–34.
- Bebchuk, L.A. 2003. *Why firms adopt antitakeover arrangements*. National Bureau of Economic Research.
- Bethel, J.E., J.P. Liebeskind, and T. Opler. 1998. Block share purchases and corporate performance. *The Journal of Finance* 53: 605–634.
- Bhagat, S., J.A. Brickley, and J.L. Coles. 1987. Managerial indemnification and liability insurance: The effect on shareholder wealth. *Journal of Risk and Insurance* 54: 721–736.
- Brickley, J.A., J.L. Coles, and R.L. Terry. 1994. Outside directors and the adoption of poison pills. *Journal of Financial Economics* 35: 371–390.
- Bushman, R.M., J.D. Piotroski, and A.J. Smith. 2004. What determines corporate transparency? *Journal of Accounting Research* 42: 207–252.
- Carline, N., S. Linn, and P. Yadav. 2002. *The impact of firm-specific and deal-specific factors on the real gains in corporate mergers and acquisitions: An empirical analysis*. Working Paper, University of Oklahoma.
- Chen, C.J.P., and B. Jaggi. 2000. Association between independent non-executive directors, family control and financial disclosures in Hong Kong. *Journal of Accounting and Public Policy* 19: 285–310.
- Chung, R., M. Firth, and J.B. Kim. 2002. Institutional monitoring and opportunistic earnings management. *Journal of Corporate Finance* 8: 29–48.
- Corporate Governance Code Monitoring Committee. 2008. *The Dutch Corporate Governance Code. Principles of good corporate governance and best practice provisions*. Hague.
- Daines, R., and M. Klausner. 2001. Do IPO charters maximize firm value? Antitakeover protection in IPOs. *Journal of Law, Economics, and Organization* 17: 83–120.

- Danielson, M.G., and J.M. Karpoff. 2006. Do pills poison operating performance? *Journal of Corporate Finance* 12: 536–559.
- Davis, G.F., and E.H. Kim. 2007. Business ties and proxy voting by mutual funds. *Journal of Financial Economics* 85: 552–570.
- Diamond, D.W. 1984. Financial intermediation and delegated monitoring. *The Review of Economic Studies* 51: 393–414.
- Eisenberg, T., S. Sundgren, and M.T. Wells. 1998. Larger board size and decreasing firm value in small firms¹. *Journal of Financial Economics* 48: 35–54.
- Fama, E., and M. Jensen. 1983. Separation of ownership and control. *The Journal of Law & Economics* 26: 301–325.
- Fosberg, R.H. 1989. Outside directors and managerial monitoring. *Akron Business and Economic Review* 20: 24–33.
- Government Commission. 2009. “German corporate governance code.” <http://www.corporate-governance-code.de>
- Hermalin, B.E., and M.S. Weisbach. 1991. The effects of board composition and direct incentives on firm performance. *Financial Management* 20: 101–112.
- . 2003. Boards of directors as an endogenously determined institution: A survey of the economic literature. *Economic Policy Review* 9: 96–118.
- Jensen, M.C. 1993. The modern industrial revolution, exit, and the failure of internal control systems. *Journal of Finance* 48: 830–879.
- Lambert, R.A., and D.F. Larcker. 1985. Golden parachutes, executive decision-making, and shareholder wealth. *Journal of Accounting and Economics* 7: 179–203.
- Lipton, M., and J.W. Lorsch. 1992. Modest proposal for improved corporate governance, a. *Business Lawyer* 48: 59–77.
- Mathiesen, H. 2002. *Managerial ownership and financial performance*. Denmark: Dissertation, Copenhagen Business School.
- Morck, R., A. Shleifer, and R.W. Vishny. 1988. Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics* 20: 293–315.
- Office of Economic and Corporate Development. 1999. *OECD principles of corporate governance*. Paris.
- Pound, J. 1988. Proxy contests and the efficiency of shareholder oversight. *Journal of Financial Economics* 20: 237–265.
- Rosenstein, S., and J.G. Wyatt. 1990. Outside directors, board independence, and shareholder wealth. *Journal of Financial Economics* 26: 175–191.
- Shleifer, A., and R.W. Vishny. 1997. The limits of arbitrage. *Journal of Finance* 52: 35–55.
- Stapledon, G.P. 1996. *Institutional shareholders and corporate governance*. USA: Oxford University Press.
- The Financial Reporting Council. 2010. *The UK corporate governance code*. London.
- Van Ees, H., T.J.B.M. Postma, and E. Sterken. 2003. Board characteristics and corporate performance in the Netherlands. *Eastern Economic Journal* 29: 41–58.
- Yermack, D. 1996. Higher market valuation of companies with a small board of directors. *Journal of Financial Economics* 40: 185–211.



James Winder

Abstract

International banks have played a key role in the financial integration of global financial markets and the economic integration of countries. Their role as geographic intermediaries has helped send capital to countries and regions that previously had difficulty attracting funds. But because of poor risk and liquidity management, the banks also played a central role in the 2008–2009 financial crisis. European banks remain weakened by the Euro-Zone’s sovereign debt crisis, as policymakers search for a resolution. This chapter presents a broad look at international banking since the 1960s, and, to a lesser extent, the role of financial institutions generally in world financial markets.

Keywords

Hedge funds · Hedge fund history · Hedge fund strategies · Hedge fund trends

53.1 Introduction

International banks have played a key role in the financial integration of global financial markets and the economic integration of countries. Their role as geographic intermediaries has helped send capital to countries and regions that previously had difficulty attracting funds. But because of poor risk and liquidity management, the banks also played a central role in the 2008–2009 financial crisis. European banks remain weakened by the Euro-Zone’s sovereign debt crisis, as policymakers search for a resolution. This chapter presents a broad look at international banking since the 1960s, and, to a lesser extent, the role of financial institutions generally in world financial markets.

J. Winder (✉)

Rutgers University, New Brunswick, NJ, USA

e-mail: jpwinder@business.rutgers.edu

This chapter comprises an introduction and three sections. Section 53.2 will discuss a brief history of the development of international banking activity in the United States, Europe, and Japan. Section 53.3 will report on historical trends in banking activity beginning in the late 1970s, and it will also assess how the financial crisis of 2008–2009 affected international banking activity. Section 53.4 will cover international banking regulation. This section will include a history of international banking regulatory efforts, which began in the 1970s. The most recent changes to the regulatory structure are defined in Basel III, which is the global response to the regulatory failures revealed during the “Great Recession.” This section will also highlight some of the debate about this new regulatory architecture designed by the Basel Committee on Banking Supervision.

53.2 Aspects of International Banking History Since 1960

International banking grew in the 1960s and 1970s in the US, Japan, and Europe for some similar reasons across the three regions. Commercial banks faced slowing domestic growth and the declining profitability of domestic banking business, and, so, they looked for greater opportunities in foreign markets. There are, however, differences, which are discussed below.

53.2.1 Japan

Burton and Saelens (1983) describes the expansion of international banking in Japan. Pushed by increasing trade with the rest of the world, Japan’s Ministry of Finance, beginning in the 1960s, began to allow greater expansion of Japanese banks in foreign markets. These developments marked Japan’s ascendance as a global economic power. Initially, the expanding foreign banking was mostly lending to businesses, but from the mid 1970s onward, Japan’s international banking presence expanded into consumer retail banking as well.

Along side the expansion of international banking that came from increasing trade, domestic economic conditions also played a role. Economic growth was quite rapid from the mid 1950s until the early 1970s. But following the first oil price shock in 1973, the trend rate of economic growth in Japan slowed, reducing the earning power of the banks. A less optimistic forecast for growth of the domestic banking market led the largest banks (the city banks) to expand their international operations. Of course, global economic growth also slowed following the first oil price shock. Global banking activity was also diminished by the failure of the German Herstatt Bank in 1974 and the second oil price shock in 1979. In order to gain market share in international banking at such a difficult time, Japan banks pushed international expansion by aggressively competing on loan terms. Commercial lending spreads offered by Japan’s banks were much narrower than spreads on comparable loans offered by US and European banks. The strategy was

successful. By the early 1980s, Japanese banks ranked second in terms of international lending to US banks.

53.2.2 The United States

Brimmer and Dahl (1975) assesses some of the special characteristics of the US expansion into international banking. The expansion of US banks into international banking needs to be viewed against the backdrop of the Bretton Woods currency agreement that prevailed from just after World War II until the early to mid 1970s. Japan's Ministry of Finance supported the growth of international banking. The US took a different approach. US policies were aimed at regulating and managing commercial bank lending to foreigners as a tool to reduce the widening balance of payments deficit. The persistent widening of the US international payments deficit in the early 1970s made sustaining the fixed rate currency system, with the dollar as its foundation, more difficult. As part of this program to manage banking sector foreign lending and investments, the US had from the early 1960s to 1974 a program of voluntary foreign credit restraint (VFCR) managed by the Federal Reserve and an interest equalization tax, which was intended to reduce the returns from investing in foreign markets. Despite efforts by President Johnson to tighten the restrictions on private foreign financial flows, the program did not lead to a narrowing of the US balance of payments deficit. The policies, perhaps predictably, had the reverse effect of encouraging the off-shore growth of US banks.

Following the removal of VFCR limits in 1974, there were two key issues for US policymakers (Brimmer and Dahl 1975). First, how should the rapidly expanding non-US banks be treated, given that the US lacked national banking? To address this issue, Congress passed the International Banking Act of 1978. The goal of the IBA of 1978 was to bring foreign banks under the same set of federal regulations as those that applied to domestic banks. There was still, however, a big advantage for foreign owned banks: they could operate in more than one state, while US banks could not.

The second issue was, were there workable policies to slow the foreign expansion of US banking corporations? Along with the VFCR and interest equalization tax, deposit reserve requirements and FDIC deposit insurance added to the cost of lending from the US. Foreign branches of US banks could avoid these costs. And, so, the result was a rapid expansion of US banking companies into London and other centers like the Cayman Islands and Bermuda in the 1960s and 1970s. US banks also had to contend with deposit rate ceilings that compromised their ability to compete for international funds. International Banking Facilities (IBFs), first proposed in 1978 and approved by the Federal Reserve in 1981, were intended to bring offshore banking activities back onshore in the US. The IBFs allowed US banks to operate free of these regulations provided they took deposits and made loans only to non-US residents. The IBFs did not represent new office locations or buildings. Rather they were a second set of books for banks that did international business.

53.2.3 Europe

Ross (1998) describes the efforts in the 1960s to promote international banking in Europe through banking clubs. The clubs were groups of large banks whose goal was to give member banks the ability to do business in other countries without having offices in foreign countries. The goal of the clubs was to slow the advance of US banks into European markets.

The banking clubs represented a business strategy for participating in international banking that had three parts. First, was the belief that the clubs would foster increased European economic integration. In September 1957 Belgium, France, Germany, Italy, Luxembourg, and the Netherlands signed the Treaty of Rome forming the European Economic Community. The Treaty made a firm commitment to the principle of competition, including the free movement of labor and free trade in services (Dutta 2007). It was hoped the banking clubs would build on this foundation and further enhance the process of European economic integration. The second reason for the founding of banking clubs was to meet the challenges posed by a growing American business presence in Europe. Between 1962 and 1969 deposits at American banks in London increased by almost 20 times, and by 1972 they had captured about 10% of all loans to UK residents (Ross 1998). American banks and American businesses more generally were seen not only as a threat to European capitalism, but also to European culture.

The third reason for the development of banking clubs was that clubs represented the best way to participate in opportunities offered by the development of international financial markets. There were three options for European banks to respond to the challenge of the growing American banking presence (Ross 1998). First, they could set up a branch system across countries, essentially following the American approach. Second, European banks could merge, forming multinational entities. And third, they could find partners with whom they could become increasingly integrated over time.

The first choice was difficult to achieve. Establishing a successful international branch system takes time, which European banks did not have given the superior managerial ability of the American banks. The second choice of merging to expand size and capability was difficult to achieve because of differences in tax laws and regulatory policies across countries. The third choice was the only real option. Find a group of partners to combat, over the short-term, America's forceful presence in European finance and commerce, and expect that over time the banks in the club would become more integrated. This was especially so since it was believed that economic integration of Europe would gain momentum. The idea was that by joining forces European banks could offer superior service to clients and compete successfully against US banks (Ross 1998).

The system of banking clubs was failing by the mid 1970s. There were conflicts of interest and business opportunities as club members began to compete against each other. Progress toward European economic union also dimmed during these years. The differences mentioned above in tax and regulatory policies, and also exchange controls, made progress toward an economic union slow.

53.3 Trends in International Lending 1977 to 2011

This section looks at trends in international data for commercial banks and financial institutions more broadly. The BIS international banking database begins in 1977 for many countries, but not until the mid 2000s for other nations. BIS data series for financial institutions also begin at various times. We will assess trends over two time periods: the full 1977–2011 interval, and a shorter interval covering the financial crisis of 2008–2009.

53.3.1 International Banking and the Relationship to Economic Activity

BIS data on international banking begin in 1977 for a large sample of countries, which are mostly Western Europe, the United States, Canada, and Japan. The sample expands to cover more emerging markets after 2000. Table 1 looks at the growth of international assets and liabilities for all countries and for the US, Europe, and Asia. Data series for many Asian countries do not begin until the mid-2000s.

The 5-year time periods reported are based around year end 2007, when the dollar amount of international assets and liabilities peaked prior to the near collapse of the global financial system in 4Q/2008. There were two periods of rapid expansion of international banking, 1977–1982 and 1982–1987, prior to the acceleration that

Table 1 Growth rates of international bank lending (annual rates of changes per time period)

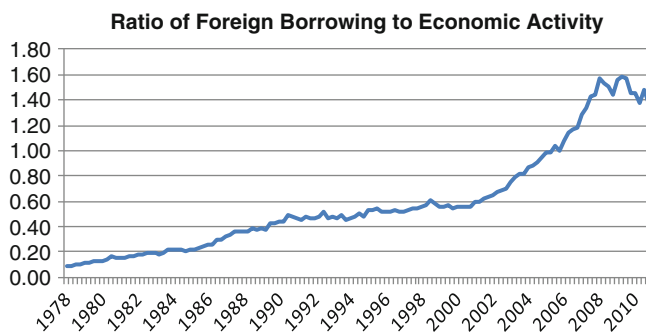
	Dec. 1977	Dec. 1982	Dec. 1987	Dec. 1992	Dec. 1997	Dec. 2002	Dec. 2007
	Dec. 1982	Dec. 1987	Dec. 1992	Dec. 1997	Dec. 2002	Dec. 2007	Mar. 2011
All countries							
Assets	20.8%	13.5%	9.3%	9.8%	7.0%	18.9%	−3.0%
Liab	20.6%	14.9%	9.2%	9.1%	6.4%	18.3%	−2.9%
United States	31.7%	6.4%	1.9%	7.4%	7.9%	18.5%	4.0%
Europe							
Assets	18.3%	9.6%	9.2%	11.2%	9.9%	19.4%	−6.3%
Liab	19.4%	9.7%	9.9%	10.5%	9.1%	18.9%	−5.3%
						Mar. 2005	Dec. 2007
						Dec. 2007	Mar. 2011
Asia							
Assets						19.3%	2.7%
Liab						16.3%	7.9%

Source: BIS

preceded the Lehman Brothers bankruptcy in September 2008. The first period corresponded to the inflationary period that resulted from the second oil price shock in 1979 and years of easy US monetary policy. There were two US recessions during this period, as the Federal Reserve under Paul Volcker shifted from targeting interest rates to targeting nonborrowed reserves. From 1982 to 1987, interest rates were generally declining, while the global economy was strengthening.

Table 1 also shows the acceleration in international lending that occurred during the credit expansion that preceded the financial crisis of 2008. From December 2002 through December 2007, international banking activity expanded at roughly similar rates across Europe, Asia, and the United States. International assets and liabilities increased from 16% to 19% per year. But, from the peak in the dollar amount of international assets and liabilities outstanding in December 2007, there has been a net decline in the global totals. Global international lending is down 3% through the first quarter of 2011. International lending in Europe declined significantly more than the global total, perhaps reflecting the large amounts of lending to troubled European governments.

We can compare the credit burst preceding the 2008 financial crisis to other periods of rapid international banking growth. Specifically, international banking activity experienced a sharp acceleration relative to economic growth beginning around 2000. Graph 1 shows international banking relative to global economic activity. The graph was created as follows. A ratio of international banking activity to global economic growth was created using BIS data on international lending and IMF data on inflation and industrial production. The IMF data cover the developed world, which produced most of the international banking activity during the entire 1977–2011 period. The international banking data are first deflated by the CPI for developed countries to create a “real” lending total, which can then be compared to production, also a real variable. An index was then created for real lending matching the base period (2005) used for the IMF industrial production index. We then take the ratio of the real lending index to the index for industrial production of the developed countries. If real lending is increasing relative to the pace of global production, then the ratio will be increasing.



Graph 1 Ratio of foreign borrowing to economic activity. (Source: BIS, IMF, author's calculations)

The graph reveals a gradually increasing trend in international banking relative to economic activity until around 2000, when a sharply increasing trend begins. Some of acceleration can be attributed to the increased number of countries participating in the BIS banking sample. But, it can be seen from Table 2 there was an acceleration in international banking activity for the group of countries covered by the banking data which begins in 1977. For example, the data for the US, Canada, France, Germany, Italy, Japan, and the UK also show an acceleration beginning in the 2002–2007 period. That is, the graph would have a similar profile if only countries whose data begin in 1977 were used to create the graph. Many versions of this graph were created. Some used international assets, some international liabilities, and some did not make any adjustment for inflation in the developed countries. The analysis always produced a graph with a similar look.

In one calculation of the lending to economic activity ratio, we used the IMF annual series for world nominal GDP, measured in US dollars, which begins in 1980. The numerator of this ratio was BIS international loans outstanding, measured in current US dollars, and the denominator was world nominal GDP. The graph has the same general shape as the graph based on global industrial production. Regardless of the computational method used, the graph has the same key feature: an intensifying use of international lending and borrowing beginning around 2000.

Some of the more intensive use of international credit reflects globalization, the expansion of world trade and financial flows relative to the size of the global

Table 2 Growth rates of international bank lending: selected countries (annual rates of changes per time period)

	Dec. 1977	Dec. 1982	Dec. 1987	Dec. 1992	Dec. 1997	Dec. 2002	Dec. 2007
	Dec. 1982	Dec. 1987	Dec. 1992	Dec. 1997	Dec. 2002	Dec. 2007	Mar. 2011
Selected countries							
Canada	17.1%	6.1%	−2.7%	13.9%	6.4%	15.8%	16.4%
France	16.7%	7.4%	13.5%	7.7%	4.9%	24.1%	−1.2%
Germany	10.3%	15.3%	14.6%	15.3%	16.1%	16.5%	−5.9%
Italy	19.1%	7.5%	10.6%	13.1%	4.8%	19.4%	−0.3%
Japan	33.5%	38.9%	10.0%	6.3%	−0.3%	14.8%	3.9%
United Kingdom	22.3%	9.1%	5.3%	11.8%	8.1%	20.1%	−6.2%
Brazil						32.4%	1.2%
Chile						3.7%	4.8%
Hong Kong			14.4%	4.8%	−9.1%	14.8%	0.7%
India						13.2%	3.0%
Mexico						10.6%	21.8%
S. Korea						13.5%	6.9%
Turkey						22.4%	−7.8%

Source: BIS

economy. There are also many current analyses, some by academics and some by financial industry researchers, which conclude the more private sector deleveraging is needed in the industrialized world before balance sheets are healthy again. It is interesting to note that our measure of international banking activity relative to global production points in the same direction. International assets and liabilities outstanding are still very high relative to the pace of global economic activity, suggesting that more global deleveraging is needed before the world's financial system is strong again.

The regional data in Table 1 and the country data in Table 2 illustrate a key point about trends in international banking activity since the end of 2007. Activity has declined in the developed countries, while it has expanded in emerging market nations. The aggregates for Europe and Asia presented in Table 1 and the individual countries reported in Table 2 show this. Some of this probably reflects the problems affecting European banks, as mentioned earlier. But, it could also reflect the growth of financial markets in IEM countries.

53.3.2 The Dominance of Developed Country Financial Institutions in International Debt Markets

Financial institutions, that is, commercial banks, investment banks, and insurance companies, from developed countries dominate the international debt market. Table 3 reports the share of total financial institution international debt outstanding issued by developed and developing nations. The data begin in 1987. Five-year intervals are again reported, while the most recent quarter 1Q/2011 is also included.

Table 3 International debt securities issued by financial institutions (by category of issuer, percent of FI total outstanding)

	Dec-87	Dec-90	Dec-95	Dec-00	Dec-05	Dec-10	Mar-11
Developed	96.2%	97.2%	94.3%	96.1%	96.6%	95.9%	95.9%
Offshore	0.5%	0.3%	0.9%	1.2%	1.0%	0.9%	0.8%
Developing	3.3%	2.5%	4.8%	2.7%	2.4%	3.2%	3.3%
Africa & ME	0.5%	0.2%	0.1%	0.1%	0.2%	0.6%	0.6%
Asia Pac	2.1%	1.4%	2.7%	1.5%	1.1%	1.3%	1.4%
Europe (IEM)	0.2%	0.5%	0.2%	0.2%	0.4%	0.7%	0.7%
Latin America	0.4%	0.3%	1.8%	0.9%	0.6%	0.7%	0.7%
Selected countries:							
BRIC	1.4%	1.2%	1.8%	0.7%	0.8%	1.3%	1.4%
MIST	1.2%	0.8%	2.2%	1.1%	0.8%	0.7%	0.7%

Source: BIS

Notes for developing countries:

Africa & ME: Africa & Middle East

Asian Pac: Asia Pacific. Excludes Japan, Australia, and New Zealand

Europe (IEM): Includes Croatia, Hungary, Poland, Russia, and Turkey

The figures show that financial institutions in developed countries had issued about 96% of total financial institution international debt outstanding in 1987 and had roughly the same market share in March 2011. This was a surprise. Given the trend in emerging market economic growth relative to economic growth in the developed world, we had thought that there would be greater penetration of emerging market financial institutions over the past 25-year period. But, the data just reinforce the notion that acquiring access to global debt markets is difficult to achieve for most emerging countries. Gaining access to global capital markets is a process that takes many years, as described in Eiteman (2010).

International debt outstanding issued by Asia Pacific financial institutions actually declined slightly from 1987 through early 2011 as a share of the total international debt obligations of financial institutions. BRIC countries did not gain market share in international debt markets, and the newly created “MIST” group of countries comprising Mexico, Indonesia, South Korea, and Turkey had a declining share. The developing region with the most success gaining access to global debt markets is Eastern Europe, Russia, and Turkey. Latin America financial institutions also had some modest success penetrating global debt markets.

53.3.3 International Money Market Instruments and the Currency of Issue

Global financial institutions have come to dominate the market for international short-term debt, most of which is commercial paper. Over 90% of international debt instruments with original maturities of less than 1 year were issued by financial institutions as of March 2011, which is up significantly from 38% in December 1990.

Table 4 begins at the end of 1990. The international money market instruments issued in what is called “euro” represented 10% of total issues outstanding. This is, of course, 8 years before the introduction of the euro on January 1, 1999. The figures before January 1, 1999 represent international money market debt issued by the

Table 4 International money market instruments: long-term history (distribution by currency of issue)

	Dec-90	Dec-95	Dec-98	Dec-99	Dec-00	Dec-05	Dec-10	Mar-11
Euro	10.1%	19.6%	17.3%	33.7%	32.2%	43.6%	41.3%	40.9%
Dollar	77.7%	59.2%	58.1%	43.4%	43.5%	29.4%	37.2%	36.9%
Pound	2.3%	5.8%	7.2%	7.3%	7.6%	18.0%	13.7%	14.9%
Swiss		5.8%	3.9%	2.0%	1.8%	2.5%	1.6%	1.4%
Yen	1.0%	4.8%	8.1%	9.5%	11.1%	1.7%	2.3%	1.9%
Financial institutions ^a	38.2%	77.2%	82.0%	86.4%	83.4%	91.8%	91.8%	92.5%

Source: BIS

^aShare of total international money market debt outstanding

countries in the Euro-Zone as of its starting date. This table also includes data for year end 1998 and for year end 1999. In just its first year, the euro captured a significant share, almost one-third of the international short-term debt market. The most recent data put the euro as the currency of issue for about 41% of international short-term debt. The dollar was the currency of issue for almost 78% of total international money market debt outstanding when the data begin in 1990. Figures for 1Q/2011 show the dollar share to be about 37%. Interestingly, the share of international short-term debt issued in yen increased substantially during the 1990s before falling sharply in the past 10 years. On a trend basis, the share denominated in pounds has risen since the data begin and since the introduction of the euro. The rise of the euro and the pound have, thus, come at the expense of the dollar and the yen.

While the long-term data show a rising trend for euro-denominated international money market liabilities, and a declining trend for dollar-denominated liabilities, data during the 2008–2009 financial crisis and beyond reveal a different story. The share of international money market debt issued in dollars increased. The figures in Table 5 begin in June 2008, the quarter before the financial crisis started. The figures also cover Greece's sovereign debt problems in the spring of 2010 and Ireland's crisis in the fourth quarter of 2010.

The share of international money market obligations issued in euro declined at year end 2008 as the financial crisis reached its most intense period in January 2009 when the Merrill Lynch/Bank of America merger almost failed. Once the depth of the crisis was reached in the first quarter of 2009 and global equity markets recovered sharply, the dominance of the euro over the dollar reasserted itself, with the euro reaching 51% of total international money market debt outstanding in September 2009. Since this peak, the euro share fell to just 41% in March, 2011, reflecting to some extent the Euro-Zone sovereign debt crisis that is still unresolved.

What is interesting is that since the Euro-Zone's sovereign debt problems first surfaced in Greece in the second quarter of 2010, the share of dollar-denominated money market debt has been steady at around 36%. The decline in the euro share

Table 5 International money market instruments: 2008 financial crisis history (distribution by currency of issue)

	Jun-08	Dec-08	Jun-09	Sep-09	Dec-09	Jun-10	Dec-10	Mar-11
Euro	47.0%	45.2%	48.3%	51.0%	47.5%	45.1%	41.3%	40.9%
Dollar	29.2%	32.7%	30.1%	29.5%	34.3%	36.4%	37.2%	36.9%
Pound	13.4%	13.1%	13.7%	12.1%	10.6%	10.5%	13.7%	14.9%
Swiss	2.1%	2.1%	2.0%	2.1%	2.2%	2.1%	1.6%	1.4%
Yen	3.6%	2.9%	2.1%	1.6%	1.8%	2.3%	2.3%	1.9%
Financial institutions ^a	91.6%	85.4%	87.8%	89.6%	91.9%	91.5%	92.5%	91.6%

Source: BIS

^aShare of total international money market debt outstanding

from 45% to 41% has been offset by an increase in the amount of international money market debt denominated in sterling from 11% to 15%.

53.3.4 International Notes and Bonds: By Type and Currency of Issue

Table 6 shows the amount of international note and bond issuance due to the activities of financial institutions. The BIS data begin in 1993. The table shows figures representing floating rate debt, straight debt, and equity-linked debt, which is primarily convertible securities.

The data show that international notes and bonds issued by financial institutions comprised 41% (line 4) of the total international long-term debt outstanding when the data begin in December 1993. Once again, financial institutions came to dominate these markets over the years, reaching almost 75% of total international long-term debt outstanding in March 2011. These figures and those for short-term debt demonstrate that international financial institutions are the dominant players in the international debt securities markets.

The share of all three types of international debt instruments increased for financial institutions during the period covered by BIS data. Financial institutions dominate the markets for floating rate and straight international debt securities. The share of equity-linked debt also increased, but the largest issuer is still corporations.

The share of international notes and bonds denominated in dollars increased from 41% in 1993 to over 50% during the years to 2000. This period included the tech bubble and one of the longest periods of sustained strong real US GDP growth and low inflation on record. Over the subsequent decade the dominance of the dollar gave way to increased issuance denominated in euros. The most recent data, for 1Q/2011, show that 45% of international notes and bonds outstanding are denominated in euros versus 38% for the dollar.

How did these trends hold up during the financial crisis of 2008 and the ensuing sovereign debt crises of Europe? Table 7 reveals some of the answer. This table

Table 6 International bonds and notes: long-term history (financial Institutions as % of total outstanding)

	Dec-93	Dec-95	Dec-98	Dec-99	Dec-00	Dec-05	Dec-10	Mar-11
Floating rate	72.6%	72.6%	86.3%	89.7%	89.5%	94.2%	95.1%	94.8%
Straight	37.8%	41.9%	52.4%	56.9%	59.9%	68.7%	66.6%	66.2%
Equity linked	24.6%	32.9%	47.2%	46.5%	47.9%	47.0%	45.9%	46.2%
All types	41.4%	46.1%	59.4%	63.9%	66.7%	75.4%	74.6%	74.2%
Memo: ^a								
Dollar	41.3%	37.7%	46.8%	48.4%	50.6%	38.6%	39.2%	38.4%
Euro	24.9%	27.7%	27.2%	28.5%	29.5%	45.2%	44.1%	45.3%

Source: BIS

^aDistribution by currency of issue as a percent of total international bonds and notes outstanding

Table 7 International bonds and notes: 2008 financial crisis history (financial institutions as % of total outstanding)

	Jun-08	Dec-08	Jun-09	Sep-09	Dec-09	Jun-10	Dec-10	Mar-11
Floating rate	95.2%	95.7%	95.7%	95.6%	95.7%	95.5%	95.1%	94.8%
Straight	70.4%	70.4%	68.9%	68.7%	68.3%	67.3%	66.6%	66.2%
Equity linked	52.8%	51.4%	49.3%	48.3%	47.7%	46.4%	45.9%	46.2%
All types	78.4%	78.9%	77.6%	77.2%	76.8%	75.4%	74.6%	74.2%
Memo: ^a								
Dollar	34.0%	36.2%	36.1%	35.5%	36.2%	40.0%	39.2%	38.4%
Euro	49.7%	47.9%	47.4%	48.1%	47.5%	43.4%	44.1%	45.3%

Source: BIS

^aDistribution by currency of issue as a percent of total international bonds and notes outstanding

begins in June 2008, just as did the table for international money market debt. The dollar gained share during the crisis period, while the euro lost some market share, just as occurred for international money market instruments.

53.3.5 The 2008–2009 Financial and European Sovereign Debt Crises

The world has suffered through two major financial crises since the autumn of 2008. The 2008 financial crisis is still affecting the economies of developed nations because private sector balance sheets still have too much leverage. And the Euro-Zone has yet to find a credible program to resolve the debt problems of several governments. The direct exposure of European banks to these governments and the indirect exposure of US banks, through interbank loans to European banks, have led to the underperformance of the financial stocks relative to broad market indexes.

To what extent have these crises affecting developed Western nations led to contagion in the banking sectors of emerging market countries? The answer is “not much” according to IMF banking data. Series of bank liabilities to non-residents, excluding liabilities to non-residents on the balance sheets of central banks, were created for emerging countries that have previously experienced banking or currency crises. The group includes Mexico, Brazil, Argentina, Chile, and Peru in Latin America. And, for Asia, this sample includes Indonesia, South Korea, Malaysia, and Thailand. The figures begin in 2007, just prior to the onset of the financial crisis (Table 8).

The constructed series show that during the worst of the financial crisis, the end of 2Q/2008 through 1Q/2009 period, there was indeed capital flight from the banking systems of most of the countries in this group. Brazil, Argentina, Chile, Peru, Indonesia, South Korea, and Malaysia all experienced net outflows of funds owed to non-residents, as might be expected in a severe financial crisis. For many of these

Table 8 Private banking system: liabilities to non-residents (selected countries, in billions of national currency)

	Mexico	Brazil	Argentina	Chile	Peru	Indonesia	S. Korea	Malaysia	Thailand
Q1 2007	1,095	265	17	6,215	4	154,076	167,373	161	1,513
Q2 2007	1,018	311	17	7,753	7	159,157	168,820	192	1,380
Q3 2007	981	312	18	9,446	8	173,499	166,772	221	1,145
Q4 2007	907	294	19	9,785	14	196,236	157,422	223	1,371
Q1 2008	1,412	294	20	9,820	17	195,410	141,936	257	1,254
Q2 2008	1,196	296	20	13,392	19	198,540	174,078	289	1,233
Q3 2008	1,020	356	20	13,833	19	219,858	133,752	278	1,176
Q4 2008	1,057	366	18	15,004	18	238,262	133,874	206	1,166
Q1 2009	1,084	336	18	11,784	16	218,092	129,780	196	1,211
Q2 2009	918	305	17	10,723	15	193,717	175,505	195	1,181
Q3 2009	1,277	285	15	14,625	13	218,925	212,019	198	1,299
Q4 2009	1,198	240	14	15,769	15	225,137	206,249	194	1,552
Q1 2010	1,322	285	13	15,274	17	260,883	218,453	193	1,615
Q2 2010	1,853	341	12	16,363	17	255,959	222,282	207	1,702
Q3 2010	2,085	357	13	16,763	21	270,622	247,931	216	1,940
Q4 2010	2,193	352	12	16,388	22	301,229	244,061	216	2,363
Q1 2011	2,344	n/a	13	n/a	27	256,705	233,423	262	2,840

Source: IMF

countries, the outflows continued for 3 months after global equity markets began to recover in March 2009.

It is interesting to compare the fund flows from the 2008 financial crisis with the fund flows during the European sovereign debt crisis, which began in the spring of 2010, when Greece was close to defaulting on its sovereign debt. From the end of 1Q/2010 through 1Q/2011, almost every country experienced net inflows into its private banking system. Indonesia is the only exception. But smoothed data show that private banking system liabilities to non-residents in Indonesia were, on average, higher in the year after the Greek crisis began to unfold than they were in the year before the crisis began. Clearly, investors view Europe's problems as not having a significant effect on the banking systems of emerging market countries.

Looking at the entire period from the end of 2Q/2008 through the end of 1Q/2011, the same point can be made. Seven of the nine countries in this sample had larger foreign liabilities in their private banking systems, suggesting that credit concerns from the Great Recession and the European debt crisis did not have lasting impacts on emerging market banking systems.

53.3.6 Balance Sheet Ratios for Global Banks

Regulatory authorities reacted to the financial crisis of 2008 with new proposals to regulate the global banking system. These new regulations, known as Basel III, will be discussed in the next section. For now, we set the stage for an assessment of the new regulatory architecture by focusing on some key measures of capital adequacy before, during, and after the Great Recession. The figures come from Bloomberg.

Table 9 reports several measures of capital adequacy for a sample of ten global banks: the five largest US commercial banks and the five largest European commercial banks. The time period covered is from year end 2000 to year end 2010. Three measures of capital adequacy are reported: (1) the tangible common equity ratio, (2) the BIS Tier 1 capital ratio, and (3) the BIS total risk-based capital ratio.

The ratios and their components are defined as follows:

- Tangible common equity capital ratio: tangible book value as a percent of tangible assets.
- Tier 1 capital: this category includes the book value of common stock, retained earnings, and non-callable, non-cumulative preferred shares.
- Tier 2 capital: this category includes subordinated debt, loan loss reserves, and non-callable cumulative preferred shares.
- Risk-Weighted Assets: Weights determined by an asset's credit risk.
- (See Saunders and Cornett 2011 for more detailed definitions.)

The Tier 1 capital adequacy ratio is defined as Tier 1 capital divided by risk-weighted assets, and, through the period covered in the table, this ratio must be at least 4%. The total regulatory capital ratio is defined as Tier 1 plus Tier 2 capital divided by risk-weighted assets, and, through the period covered in the table, this

Table 9 Capital adequacy ratios for selected global banks (percent, at year end)

	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
I. United States											
Bank of America											
Tangible common	5.86	4.50	2.83	3.35	4.14	4.24	4.79	4.91	5.83	5.96	6.12
Tier 1	11.24	10.40	9.15	6.87	8.64	8.25	8.20	7.85	8.22	8.30	7.50
Total Risk based	15.77	14.66	13.00	11.02	11.88	11.08	11.73	11.87	12.43	12.67	11.04
JPM Chase											
Tangible common	5.59	5.27	3.83	4.76	4.87	4.82	5.09	4.59	4.22	3.65	3.57
Tier 1	12.10	11.10	10.90	8.40	8.70	8.50	8.70	8.50	8.24	8.29	8.50
Total risk based	15.50	14.80	14.80	12.60	12.30	12.00	12.20	11.80	11.95	11.88	12.00
Citigroup											
Tangible common	6.89	6.49	1.56	3.11	3.79	4.68	4.51	4.53	4.70	4.98	5.52
Tier 1	12.91	11.67	11.92	7.12	8.59	8.79	8.74	8.91	8.47	8.42	8.38
Total risk based	16.59	15.25	15.70	10.70	11.65	12.02	11.85	12.04	11.25	10.92	11.23
Wells Fargo											
Tangible common	6.72	5.44	2.33	5.93	7.15	6.12	6.25	6.11	5.71	5.59	5.93
Tier 1	11.16	9.25	7.84	7.59	8.95	8.26	8.41	8.42	7.60	6.99	7.29
Total risk based	15.01	13.26	11.83	10.68	12.50	11.64	12.07	12.21	11.31	10.45	10.43
US Bancorp											
Tangible common	5.22	5.10	3.19	4.74	5.17	6.29	6.45	6.12	5.71	5.52	6.36
Tier 1	10.50	9.60	10.60	8.30	8.80	8.20	8.60	9.20	8.00	7.70	7.10
Total risk based	13.30	12.90	14.30	12.20	12.60	12.50	13.10	13.60	12.40	11.70	10.90

(continued)

Table 9 (continued)

	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
II. Europe											
BNP Paribas											
Tangible common	2.66	2.36	1.97	2.09	2.33	2.32	2.49	2.73	2.65	2.36	2.71
Tier 1	11.40	10.10	7.80	7.30	7.40	7.60	7.50	9.40	8.10	7.30	7.10
Total risk based	14.50	14.20	11.10	10.00	10.50	11.00	10.10	12.90	10.90	10.60	10.10
RBS Group											
Tangible common	4.22	3.57	1.63	0.17	2.50	2.05	2.58	2.28	2.72	2.52	2.27
Tier 1	12.90	14.10	10.00	7.30	7.50	7.60	7.00	7.40	7.30	7.10	6.90
Total risk based	14.00	16.10	14.10	11.20	11.70	11.70	11.70	11.80	11.70	11.50	11.50
Barclays											
Tangible common	2.85	2.81	1.28	1.23	1.26	1.11	2.10	2.75	2.82	2.95	2.86
Tier 1	13.50	13.00	8.60	7.80	7.70	7.00	7.60	7.90	8.20	7.80	7.20
Total risk based	16.90	16.60	13.60	12.10	11.70	11.30	11.50	12.80	12.80	12.50	11.00
Deutsche Bank											
Tangible common	1.76	1.78	0.95	1.38	1.53	2.20	2.56	2.70	3.44	3.72	1.78
Tier 1	12.30	12.60	10.10	8.60	8.50	8.70	8.60	10.00	9.60	8.10	7.80
Total risk based	14.10	13.90	12.20	11.60	12.50	13.50	13.20	13.90	12.60	12.10	13.10
HSBC											
Tangible common	4.74	4.10	2.65	3.72	3.89	4.03	4.10	4.56	4.66	4.67	4.63
Tier 1	12.10	10.80	8.30	8.70	9.40	9.00	8.90	8.90	9.00	9.00	9.00
Total risk based	15.20	13.70	11.40	11.40	13.50	12.80	12.00	12.00	13.30	13.00	13.30

Source: [Bloomberg](#)

ratio must be at least 8%. The global regulatory structure is covered in detail in Sect. 53.4.

The figures in the table make an important point. This group of large banks satisfied the Tier 1 and total risk-based capital standards throughout the 2000–2010 period. This was especially true during the crisis years of 2008 and 2009. None of the banks in this sample missed either of the BIS capital standards when the financial crisis was at its most intense point at the end of 2008.

The tangible common equity ratio (TCER), which was not included in either the Basel I or II accords, reveals a banking system under considerably more stress than portrayed by the Tier 1 and total risk-based capital ratios. It is fair to say that the European banks in our group had lower TCERs than did the group of US banks. Of the largest banks, only one US bank had a TCER below 2% at the end of 2008, while only one of Europe's largest banks had a TCER above 2%.

53.4 Basel III and the Regulatory Response to the Financial Crisis of 2008

This section traces the history of international banking regulation from the formation of the Basel Committee on Banking Supervision in the 1970s through Basel III. This section also includes summaries of key criticisms of Basel III.

53.4.1 A Brief History of International Banking Regulation

Jablecki (2009) relates the beginning of a determined effort to create an international banking regulatory structure to the 1974 collapse of two international banks: Franklin National in the US and Bankhaus Herstatt in Germany. A change of controlling interest in Franklin National from an American industrialist to an Italian financier led the bank into significant foreign currency speculation (among other risky practices) that ultimately led to insolvency. The Herstatt failure highlighted, as never before, counterparty/settlement risk. Forced into liquidation by German regulators, the bank failed to make US dollar payments after receiving deutschemarks from clients. The international nature of the problems that ultimately brought down both banks and threatened the global financial system focused policymakers on the growing size of cross-border capital flows, and the increased integration of global financial markets. Banks whose domestic activities might be adequately supervised by national authorities, might have international businesses that were inadequately regulated, potentially threatening the stability of the global financial system.

Recognizing the potential threats posed by the increasing size and interconnectedness of global financial transactions, the heads of the G-10 central banks established the Basel Committee on Banking Supervision (BCBS) in 1974 under the auspices of the Bank for International Settlements (BIS). The BCBS was then and remains the key policy developing entity for the global financial regulatory effort.

Khan (2002) takes a more endogenous view of the origins of international financial regulation. He focuses on deregulation, technology, and financial innovation as the primary forces behind the push to develop a global financial architecture in the 1970s and 1980s. First, the lifting of deposit rate ceilings in the US and the collapse Bretton Woods fixed exchange rate system occurred during a period of high and unpredictable inflation. This unstable economic environment created a demand for hedging tools, possibly denominated in non-dollar currencies. Second, technology and communication advances blurred the long prevailing barriers between banks and nonbanks, which intensified competitive pressures across the whole of the global financial industry. Third, the globalization of commercial banking led to competition between domestic and international banks, sparking debates about differences in regulations among banks across national borders.

The first action taken by the BCBS was in 1975. The Basel Concordat proposed that the country in which a bank is doing business, and not the bank's home country, would be the regulatory authority for liquidity and solvency issues. The activities of foreign banking subsidiaries would be regulated by the host country and not the parent's home banking authority (Jablecki 2009).

The next policy from the BCBS was the Basel Accord of 1988, otherwise known as Basel I. Basel I established the internationally recognized capital adequacy standards that are reflected in the Tier 1 and total risk-based capital ratios presented in Table 9 above. The G-10 governments quickly adopted these standards, and by the late 1990s Basel I had been implemented in over 100 countries (Jablecki 2009). The impetus for global capital adequacy ratios reflected concerns that national regulators would be tempted to relax capital standards during periods of financial stress, thus weakening the balance sheets of domestic banks. To remove this temptation, global standards were necessary. The two capital adequacy ratios would act as constraints on the expansion of bank credit. It was not the monetary authority that determined how fast bank credit could expand, it was the capital adequacy rules imposed by regulators (Disyatat 2008).

Basel I was the first step on the road to a coordinated regulatory structure that could strengthen the banking systems of the industrialized nations. But, it did have deficiencies. First, it dealt with credit risk, but not other types of risk, such as market, interest rate, or operational risks. So banks had an incentive to perhaps take more market and interest rate risk than they would otherwise take. And because the Accord focused only on credit risk, regulators could not be confident that banking stability was enhanced. Second, it gave banks an incentive to economize on capital. That is, banks can increase their total capital ratio by either acquiring more capital or by restructuring their balance sheets to reduce risk-adjusted assets. One way this occurred was through asset securitization. And because it is easier to securitize higher quality assets, banks kept on their books the riskiest assets in any particular risk category (Balin 2008). Also, many banks came to believe that if they satisfied the Tier 1 and total risk-based capital standards, they had done all that was needed to achieve stability (Balin 2008). It is clear from the data in Table 9 and the events of 2008 that just achieving, and even exceeding, the required BIS capital adequacy percentages does not eliminate the risk of a bank facing extreme financial stress.

Another important consequence of the Basel I risk-based capital rules was the procyclicality of bank capital. A 1999 BIS study covering G-10 countries concluded that bank capital is sensitive to the business cycle. Recessions tend to diminish bank capital, and so contribute to a reduction in bank credit just when new lending is needed. And bank capital tends to increase in expansions, tending to increase credit expansion, just when a slowing in lending is needed to maintain macroeconomic balance (BIS 2011). This point is specifically addressed in Basel III, which is discussed below.

The deficiencies of Basel I sent the Basel Committee back to work to accomplish two goals. First, broaden the earlier rules so national regulators would focus on other aspects of banking risk, not just credit risk. And, second, make the asset risk-weighting categories more accurate, so securitization did not leave balance sheets riskier. For example, under Basel I corporate loans had a risk weight of 100%, regardless of whether the borrower was a AAA credit or a BBB credit. Under Basel II, weaker corporate borrowers have higher risk weightings (Saunders and Cornett 2011).

In 1999, the BCBS put forward a three-pillar approach to regulating banks, which was known as Basel II. Pillar 1 is Regulation, that is, the capital adequacy requirements set by regulators. Pillar 2 is Supervision, which covers enforcing regulations and monitoring the health of banks on an ongoing basis through regular contact between supervisors and bank management. Pillar 3 is Market Discipline, which promotes the disclosure of information so markets can accurately price a bank's risk, and, therefore, enforce good behavior. Basel II shifted the emphasis of bank regulation away from capital adequacy ratios and toward supervision and market discipline (Khan 2002). In the aftermath of the 2008–2009 financial crisis, regulators returned to quantitative rules as the key safeguards of banking system stability.

53.4.2 The Failure of Basel II to Prevent Financial Crisis

While the Tier 1 and total risk-based capital ratios were important steps in standardizing global capital adequacy ratios, and the three pillars of Basel II broadened the scope of bank supervision, they did not prevent, nor did they signal, the near collapse of the global financial system that occurred in late 2008. This point is illustrated by the figures in Table 9, as discussed above. None of the banks in the sample failed to meet the Tier 1 or total capital ratio standards for year end 2008, when the crisis was most intense. Yet, many US financial institutions required central bank and federal government support to survive. According to documents released by the Federal Reserve and analyzed by Bloomberg, the Fed secretly lent about \$1.2 trillion to US and European financial institutions in late 2008 to prevent a global systemic failure (Keoun and Kuntz 2011).

Identifying the causes of the “Great Recession” provides an understanding of the motivations behind Basel III. Economics Europe (2010) lists several causes of the financial crisis.

- (a) Regulatory policies were inadequate, particularly with regard to risk management.
- (b) The capital banks were required to hold was inadequate, allowing banks to build up too much leverage.
- (c) Banks had become dangerously illiquid.
- (d) Countries did not have deposits safety systems that were seen as credible by the public.
- (e) Procedures for resolving bad loans and investments were inadequate.

Other factors frequently mentioned as contributing to the crisis include:

- (f) US Federal Reserve policy was too easy, creating an environment that encouraged excessive risk taking.
- (g) Managers failed to understand the risks of funding illiquid assets in the repo market.
- (h) The risks of off-balance-sheet activities were not fully reflected in BIS capital requirements.
- (i) Stress testing of balance sheets by regulators and banks was inadequate.

This list of regulatory and managerial failures highlights the problems the BCBS addressed when the committee revised the regulatory structure for international banks.

53.4.3 Basel III

The Basel III policies emphasize quantitative standards for measuring and monitoring the financial health, both short-term and long-term, of internationally active banks. The three pillars of Basel II are still in place, but Pillar 1 has been significantly strengthened. Capital adequacy standards are toughened, and several new regulatory rules are added to improve liquidity. The intent of Basel III is clear: reduce the leverage creating potential of the international banking system, and provide for liquidity adequate to withstand both immediate and longer-term funding crises.

The major changes brought by Basel III are summarized below and in Table 10 (BIS 2011).

- (a) The Tier 1 capital minimum standard will be raised from the current 4% to 6% of risk-weighted assets by January 1, 2015. Within this requirement, Common Equity Tier 1 must be at least 4.5% of risk-weighted assets at all times when the transition period ends in 2015.
- (b) A capital conservation buffer is imposed, increasing from 0.625% on January 1, 2016 to 2.5% as of January 1, 2019. The purpose of the capital conservation buffer is to “ensure that banks build up capital buffers outside times of periods of stress which can be drawn down as losses are incurred.” (BIS 2011, page 54).

Table 10 Summary of Basel III requirements (all dates as of January 1)

	2011	2012	2013	2014	2015	2016	2017	2018	2019
Min. common equity ratio			3.5%	4.0%	4.5%	4.5%	4.5%	4.5%	4.5%
Capital conservation buffer						0.625%	1.250%	1.875%	2.5%
Min. common equity plus			3.5%	4.0%	4.5%	5.125%	5.750%	6.375%	7.0%
Capital conservation buffer									
Min. Tier-1 capital ratio			4.5%	5.5%	6.0%	6.0%	6.0%	6.0%	6.0%
Min. total capital ratio			8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%
Min. total capital plus			8.0%	8.0%	8.0%	8.625%	9.250%	9.875%	10.5%
Capital conservation buffer									
New ratios:									
Leverage					2015: disclosure begins				
Liquidity coverage	2011: observation period begins			2015: Introduce minimum standard.					
Net stable funding	2011: observation period begins			2018: introduce min. standard					

Source: BIS

The intent is to force banks to reduce dividend payments, share buy-backs, and employee bonuses when minimum capital requirements might not be satisfied.

- (c) The minimum total capital ratio remains at 8%, but the addition of the capital conservation buffer puts the total capital plus conservation buffer ratio at 10.5% when the measure is fully phased in on January 1, 2019.
- (d) Along with the capital conservation buffer, national regulators may impose a countercyclical buffer when “excess aggregate growth is judged to be associated with a build-up of system-wide risk. . .” (BIS 2011, p. 57). This buffer can vary between 0% and 2.5% of total risk-weighted assets. The countercyclical buffer is intended to mitigate the procyclical nature of bank capital. It can be reduced or removed when the business cycle slows.
- (e) A leverage ratio is introduced because many banks had excessive on- and off-balance sheet leverage in the quarters prior to the onset of the financial crisis. This enormous build-up of leverage occurred at the same time that many banks had strong risk-based capital ratios. Reducing leverage caused banks to sell assets, which drove asset prices down further, leading to declines in bank capital, and diminished credit availability, which made the contraction of economic activity more pronounced. The BCBS will test a leverage ratio based on the new definition of Tier 1 capital, and it will also test alternative ratios where the capital measures are Tier 1 common equity and total risk-based capital. The exposure measure used in the denominator is non-risk-weighted. The leverage ratio is intended to be a simple, transparent, and supplementary measure to the risk-based capital ratios. (BIS 2011, p. 61.)

Basel III also announces two minimum standards for funding liquidity.

- (f) The Liquidity Coverage Ratio (LCR) is intended to ensure that a bank can meet its liquidity needs for a 30-day interval under a scenario of significant stress. The ratio is defined as,

$$LCR = \frac{\text{stock of high-quality liquid assets}}{\text{Total net cash outflows over the next 30 calendar days}} \geq 100\%$$

This short-term liquidity ratio represents a significant change from previous regulatory rules, for banks are expected to satisfy this requirement on a continuous basis, and the stress scenario contains many of the shocks that occurred during the financial crisis. That is, banks must now maintain holdings of high quality short-term liquid assets assuming that many of the conditions that prevailed in 2008–2009 might recur. For example, the stress scenario includes: partial losses of retail and wholesale funding, a potential downgrade of up to three notches, unanticipated draw downs of unused credit lines the bank has provided to clients, increased collateral haircuts in repo markets, and the partial loss of available repo financing for some assets (BIS 2010).

- (g) The Net Stable Funding Ratio (NSFR) is intended to insure that long-term assets are funded with an adequate stream of stable liabilities over a 1-year horizon. The goal is to reduce reliance on short-term sources of funding at times when liquidity is inexpensive and readily available. And, just like with the LCR, this ratio is calculated under a stress scenario. The long-term stress scenario includes

a significant decline in profitability, a potential debt downgrade, and a material event that undermines the creditworthiness of the bank (BIS 2010). This ratio is defined below, and once again, at least 100% coverage of the need for longer-term stable funding is necessary.

$$\text{NSFR} = \frac{\text{Available amount of stable funding}}{\text{Required amount of stable funding}} \geq 100\%$$

53.5 Conclusion

Basel III imposes significant new constraints on the behavior of global banks. Indeed, the excessive leverage ratios of 2008 are now impossible to achieve given the minimum Common Equity Ratio of 4.5% that will be effective in 2015. This is all much to the satisfaction of central banks and regulators. But what are the costs of this new, heavier hand of regulation? Are there other ways to improve the supervision of global banks?

Standard and Poor's (2010) notes that the definition of high quality liquid assets in the numerator of the Liquidity Coverage Ratio is "significantly more restrictive than the standards central banks typically maintain for collateral eligibility under the liquidity facilities that serve as a key backdrop to the banking system." (S&P 2010, p. 4.) S&P also notes that the LCR assumes that all unused credit facilities are drawn down by clients, while banks are assumed not to be able to use any of their credit lines. The S&P fear is that imposing this condition might impede the flow of funding through the interbank market.

S&P (2010) also believe that the Net Stable Funding Ratio, by diminishing the traditional bank asset/liability mismatch, could diminish banks' maturity transformation role in the economy.

The Institute for International Finance has written extensively about the practices it believes will best foster financial stability. Its focus is on developing a set of best practices for banks to manage risk and increase transparency, rather than on higher capital requirements. The IIF states, "The purview of the macroprudential authority needs to extend beyond the existing regulatory perimeter. It must have access to all the information needed to assess risks to financial stability. . . ." (IIF 2011a, p. 10). And "Regulation alone cannot create a safer financial system—it can only do so if partnered with improved industry practices and a strengthened global supervisory system." (IIF 2011b, p. 11.)

The IIF believes that a strong and open relationship between banks and their supervisors is the best policy for maintaining financial stability. The IIF goes so far as to advise that banks take the initiative in providing necessary information to regulators. And banks should be willing to bear some cost, if doing so improves the quality of supervision (IIF 2011b).

Another study prepared for the City of London Corporation by Europe Economics is more critical of Basel III. For them, the key issues are the development of resolution regimes, especially those that cause bond holders to take losses and liquidate holdings,

and the maintenance of sufficient liquidity. They do not believe that a convincing case for higher capital requirements was made (Europe Economics 2010).

Much attention will be directed in coming years to the effects of Basel III on the profitability of international banks. The phase-in of many aspects of Basel III, such as the new Tier 1 capital ratio, the LCR and the NSFR, will give the BIS time to understand how the new regulatory structure affects both earnings and the volatility of earnings. Regulators will also need to assess how significantly increasing the capital and liquidity costs of lending affect the availability and terms of credit. Credit growth is necessary for economic growth throughout most of the world. Regulators will need to weigh the benefits of a more stable global financial system against the potential economic costs of diminished credit availability.

References

- Balin, Bryan J. 2008. Basel I, Basel II, and emerging markets: A nontechnical analysis. The Johns Hopkins University School of Advanced International Studies, May 2008.
- BIS. Basel III: A global regulatory framework for more resilient banks and banking systems. December 2010, revised June 2011.
- BIS. Basel III: International framework for liquidity risk measurement, standards and monitoring. December 2010.
- BIS. Long-term Issues in International Banking. Committee on the Global Financial System, Bank for International Settlements, July 2010.
- , *Quarterly Review*.
Bloomberg.
- Brimmer, Andrew F., and Frederick R. Dahl, "Growth of American International Banking: Implications for Public Policy," *The Journal of Finance*, Vol. 30, No. 2, May 1975.
- Burton, F.N., and F.H. Saelens. 1983. Trends in the Growth of Japanese International Banking in Western Europe, *Intereconomics*, July/August 1983.
- Disyatat, Piti. 2008. Monetary policy implementation: Misconceptions and their consequences. BIS working paper no. 269.
- Dutta, Manoranjan. 2007. *European Union and the Euro Revolution*. Elsevier.
- Eiteman, David K., Arthur I. Stonehill, and Michael H. Moffett. 2010. *Multinational business finance*. Prentice Hall.
- Europe Economics, Comments on the future of banking regulation. May 2010.
- IMF, International Financial Statistics Online.
- Institute of International Finance. 2011a. Macroprudential oversight: An industry perspective, Submission to the International Authorities. July 2011.
- . 2011b. Achieving effective supervision: An industry perspective. July 2011.
- Jablecki, Juliusz. 2009. The impact of Basel I capital requirements on bank behavior and the efficacy of monetary policy. *International Journal of Economic Sciences and Applied Research* 2 (1): 16–35.
- Keoun, Bradley, and Phil Kuntz. Wall Street Aristocracy Got \$1.2 Trillion in Secret Loans. *Bloomberg News*, August 22, 2011.

- Khan, Mohsin S.. 2002. New issues in bank regulation. *The Pakistan Development Review*, Winter 2002: 333–356.
- Ross, Duncan M.. European Banking Clubs in the 1960s: A Flawed Strategy. *Business and Economic History*, Volume Twenty-seven, no. 2, Winter 1998.
- Saunders, Anthony, and Marcia Millon Cornett. 2011. *Financial institutions management: A risk management approach*. 7th ed. McGraw-Hill.
- Standard & Poor's. Basel III proposals could strengthen banks' liquidity, but may have unintended consequences. *Global Credit Portal*, April 15, 2010.



Hedge Funds: Overview, Strategies, and Trends

54

John M. Longo

Abstract

Hedge funds are quite likely the most important investment vehicles in the financial markets today since they control roughly \$3.8 trillion in assets and account for the bulk of trading volume across many asset classes. They also tend to attract “the best and brightest” investors due to high levels of remuneration and fewer investment restrictions relative to public investment funds. This chapter provides its readers with a firm understanding of the hedge fund industry, their most common investment strategies, and the trends shaping the field. It also provides detailed examples of three specific hedge fund strategies: dual share class arbitrage, pairs trading, and activist investing.

Hedge funds in their various forms may play a useful role in most individual and institutional portfolios due to the prospect of higher risk adjusted returns, as well as the lack of correlation with traditional stock and bond portfolios. In short, they may improve the Markowitz Efficient Frontier of most portfolios. The industry is still evolving on a global basis and its continued growth will likely result in the further compression of fees, payment for alpha instead of beta, increased liquidity, improved transparency, wider accessibility, and increased regulation.

Keywords

Hedge funds · Hedge fund history · Hedge fund strategies · Hedge fund trends

J. M. Longo (✉)

Rutgers Business School – Newark & New Brunswick, Piscataway, NJ, USA

e-mail: jlongo@business.rutgers.edu

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,

https://doi.org/10.1007/978-3-030-91231-4_54

1249

54.1 Introduction

Hedge funds are quite likely the most important investment vehicles in the financial markets today. Their managers, such as George Soros, John Paulson, James Simons, Carl Icahn, David Tepper, and Steven Cohen, are some of the most closely watched and highly paid financial professionals in the world. Hedge funds manage roughly \$3.8 trillion in assets, a large sum of money, but small in relation to the size of mutual funds (\$24 trillion worldwide) and pension funds (\$51 trillion worldwide).¹ However, the influence of hedge funds is arguably much larger than the amount of their assets under management due to the leverage employed by most funds, as well as the active nature of many of their trading strategies. Hedge funds also account for a large share of the trading volume across many other financial instruments, such as credit default swaps.

This chapter is designed to give its readers a firm understanding of the hedge fund industry, their most common investment strategies, and the trends shaping the field. In order to accomplish these goals, we first need to discuss some background on hedge funds and how they work.

54.2 Hedge Fund Defined

There is no uniformly agreed upon definition of a hedge fund. As with defining risk, the precise definition is in the eye of the beholder. In fact, many hedge funds have no “hedge” at all and choose to increase risk. A common “working” definition of a hedge fund is that it is a private investment vehicle that charges both asset- and incentive-based fees and is sold to institutional and high net worth investors. This definition would cover the vast majority of funds that hold themselves out to prospective investors as hedge funds. However, it would also encompass most private equity funds and other partnerships.

54.3 History of Hedge Funds

Alfred Winslow (A.W.) Jones is commonly cited as the individual responsible for creating the first hedge fund, Editor (2020). His fund commenced in 1949 and lost capital in only three of the 34 years of its operation.² There is some debate on the topic of who created the first hedge fund. During a trip that I took with a group of students from Rutgers Business School to meet with Warren Buffett in 2006, he stated his belief that Benjamin Graham started the first hedge fund. Buffett noted that Graham, through his Graham – Newman Partnership, went long as well as short in his portfolio

¹The source of the number for assets under management is Preqin, while Statista is the source for mutual fund and pension fund assets. Norrestad (2021) is the source of the pension fund data.

²Value Walk is the source of Jones’ investment performance.

and charged his investors an incentive fee. There were likely other investment partnerships (e.g., commodity pools), in operation before Graham-Newman, but scant evidence exists on the details of their operations or investment strategies.

Mallaby (2010) argues that Jones' fund was likely the first to combine four characteristics that are shared by many hedge fund managers today:

1. A performance or incentive fee of 20%, derived from the same profit-sharing amount that Phoenician sea captains took from the bounty of their successful voyages.
2. The conscious avoidance of regulation. As we shall see, the regulatory screws are slowly turning on the hedge fund industry, but they are still regulated much less than other financial service firms.
3. Active short selling, in the context of a dynamic long/short investment strategy.
4. The use of leverage, also on a dynamic basis.

The original A.W. Jones Model may be expressed as follows:

$$\text{Net Exposure (\%)} = (\$ \text{ Long} - \$ \text{ Short}) / \$ \text{ Equity} \quad (1)$$

where:

\$ Long = the dollar value of long positions held by the fund

\$ Short = the dollar value of short positions held by the fund

\$ Equity = the net equity in the fund

If Jones were bullish on the direction of the overall stock market, \$Long would exceed \$Short and the fund would be considered "net long." If he were bearish, the reverse would be true and the fund would be considered "net short." If the sum of the absolute value of \$Long and \$Short, known as "gross exposure," was greater than \$Equity, then the fund would be leveraged. The earliest hedge fund managers varied each of the elements of the Jones model based on their investment outlooks.

For example, if a fund has \$100 million in investor capital and controls \$200 million in long positions and \$50 million in short positions, the fund would have a gross exposure of \$250 million and net exposure of \$150 million in dollar terms, or 150 in percentage terms.

54.4 Notable Hedge Fund Managers

The pioneering hedge funds of Jones, and likely Graham, paved the way for many other hedge fund managers who went on to earn incredible fortunes, Editor (2021). Others, such as the partners of Long Term Capital and Bernard Madoff were noteworthy for more infamous reasons. Below is a small sampling of some of the most widely discussed hedge fund managers over the past 20–30 years.

Arguably, the most famous hedge fund manager is George Soros. He became known as "The man who broke the Bank of England," after his fund, The Quantum Fund, reportedly made a billion dollars in one day (September 16, 1992) shorting the

British Pound prior to its withdrawal from the European Exchange Rate Mechanism (ERM), a precursor to the Euro. The trade was suggested by his colleague, Stanley Druckenmiller, and Soros “supersized” the position since the downside minimal and upside substantial. Soros himself soon became a billionaire, earning compensation the likes that no one had ever witnessed in the financial services field, except for perhaps Michael Milken in his peak earning years, prior to his downfall.

John Meriwether was the Managing Partner of Long-Term Capital Management, once considered the “Dream Team” of hedge fund managers. Its partners not only included Meriwether and many of his former trading colleagues at Salomon Brothers, but also Nobel Laureates (Myron Scholes and Robert Merton) and the former Vice Chairman of The Federal Reserve (David Mullins). Long Term Capital generated returns of approximately 40% per year from its 1994 launch until its downfall that began in the middle of 1998, triggered by Russia’s default on its sovereign debt. The fund at one point controlled \$100 billion in assets through an enormous and unintended leverage ratio of 250 to 1.³ The Federal Reserve organized a bailout of Long Term Capital by a consortium of Wall Street firms in order to avoid the risk of the implosion of the world financial system.

James Simons is the founder and former manager of perhaps the most successful hedge fund ever, the Renaissance Medallion Fund. A *Dealbreaker* article estimates that Medallion has averaged 39.1% *annually* since its 1988 inception through 2020.⁴ Simons and his firm, Renaissance Technologies LLC, relies heavily on mathematical models to trade a range of financial securities, including commodities, currencies, and equities. Simons no longer runs the firm on a day-to-day basis but remains its largest shareholder.

John Paulson is a hedge fund manager who was the subject of Gregory Zuckerman’s (2009) book, *The Greatest Trade Ever: The Behind-the-Scenes Story of How John Paulson Defied Wall Street and Made Financial History*. The book details Paulson’s shorting of the subprime mortgage market through the use of credit derivatives. According to *Alpha*, a popular hedge fund-oriented periodical, Paulson earned \$1 billion in 2007, \$2 billion in 2008, and \$2.3 billion in 2009. David Tepper, a well-known hedge fund manager in the distressed debt and equity area, also reportedly earned in excess of \$2 billion in 2009.⁵

Bernard Madoff is perhaps the most infamous hedge fund manager. Madoff “managed” a series of separate accounts for individuals, institutions, and funds of hedge funds. His “strategy” appeared to generate remarkably consistent returns of 8–12% per annum over a more than two-decade period, with very few losses. Of course, it was uncovered that Madoff ran the biggest Ponzi scheme ever, resulting in estimated losses of \$18 billion.⁶ On June 19, 2009, Madoff was sentenced to

³Lowenstein (2000) cites the 250:1 leverage ratio for Long-Term Capital Management.

⁴Shazar (2021) is the source of the performance of the Renaissance Medallion Fund since inception.

⁵Grant (2009) is the source of the income of Tepper and other hedge fund managers in the period around The Great Recession.

⁶CBS News (2010) is the source of the estimate of Madoff losses.

150 years in prison for his crimes. He died in prison in 2021. New hedge fund regulations, specified in the Dodd-Frank Act of 2010, were in partial response to the massive Madoff fraud.

In 2018, Warren Buffett won a high profile \$1 million bet, that eventually swelled to \$2.2 million, with a well-known fund of hedge fund manager, Jeffrey Tarrant, whose firm was Protégé Partners. Mr. Tarrant died in 2019, and his firm does not appear to be active today. The bet took place over a 10-year period with Buffett betting the S&P 500 would outperform the five handpicked funds of Tarrant. Buffett handily won the bet with the S&P 500 rising nearly 126% over the decade versus only 35% for the fund of funds.⁷

More recently, the implosion of Bill Hwang's Archegos Capital Management family office in 2021 will likely lead to increased regulation for hedge funds and family offices. Archegos was set up by Hwang, a "Tiger cub" who ran the hedge fund firm, Tiger Asia Management. In 2012, Hwang closed Tiger Asia and set up Archegos. Family offices currently have little regulatory oversight. However, Archegos lost a shocking \$20 billion in less than 2 days, engaging in heavily leveraged bets on equities across a number of brokerage firms, often through credit derivatives.⁸ Not only were the Archegos assets virtually wiped out, but its implosion resulted in billions of dollars of losses to the hedge funds that traded with Archegos.

54.5 Growth of the Hedge Fund Industry and Hedge Fund Performance

Hedge funds were somewhat of a sleepy cottage industry until a Loomis (1966) *Fortune* article, entitled "The Jones Nobody Keeps Up With," about A.W. Jones. The article noted the outstanding performance of Jones' fund. It increased 670% over the 10-year period ending May 1965. In contrast, the best performing mutual fund over the same time period, Dreyfus Fund, increased 358%. A spawn of imitators soon followed.

The enormous compensation earned by the managers of the most successful hedge funds did not go unnoticed on Wall Street. As we noted earlier, George Soros' Quantum Fund reportedly made a billion dollars in a single day in 1992. His firm's cut was likely 20% of the bounty. Star traders and portfolio managers, "the best and brightest," realized they could earn much more money running their own hedge funds than in almost any other profession *and* operate in an entrepreneurial environment with fewer investment and regulatory constraints.

Another important driving force behind the historical growth of hedge funds was the influx of institutional money. From the founding of the hedge fund industry in the first half of the twentieth century through the 1980s, high net worth investors, primarily from the USA and Europe, comprised the bulk of hedge fund clients.

⁷ Buffett's bet with Protégé Partners is detailed in Wattles (2018).

⁸ Archegos' \$20 billion loss in 2 days is discussed in Schatzker et al. (2021).

Table 1 Hedge fund performance during recent bear markets

Year	Dow Jones CS	EAFE	S&P 500
	HF Index	Int'l Index	
2000	4.85%	−14.17%	−9.10%
2001	4.42%	−21.14%	−11.90%
2002	3.04%	−15.94%	−22.10%
2008	−19.10%	−39.64%	−37.20%
2020Q1	−8.98%	−22.80%	−19.60%

Sources: Dow Jones Credit Suisse, Standard and Poors, and MSCI

Being an investor a private hedge fund with limited membership held some cachet in individual social circles. However, over the past three decades, institutional investors, and affluent investors in the emerging markets, have been responsible for the bulk of incremental hedge fund assets. For example, David Swenson, who served for decades as Chief Investment Officer of Yale's \$30+ billion endowment, was one of the first institutional investment managers to allocate a sizeable amount of capital to hedge funds. Many others in the pension fund and endowment worlds soon thereafter tried to emulate the success of the "Yale Model." Forbes' Billionaires List increasingly includes individuals from the emerging markets and some of their wealth inevitably makes its way into the hedge fund arena.

Lastly, hedge funds turned in modest nominal, but strong relative, performance during the 2000–2002 bear market in global equities, spurring a dramatic increase in asset growth. Investors, incorrectly, assumed it would be hard to lose money with most diversified hedge fund portfolios. However, aggregate hedge fund returns in the bear markets of 2008 and 2020Q1 were disappointing on a nominal basis, as many hedge funds exhibited no "hedge" at all. Hedge fund returns from April to December of 2020 were positive; hence they are omitted from the table to focus on periods of negative returns. Table 1 shows the returns for these bear market periods for The Dow Jones Credit Suisse Hedge Fund Index, as well as for the S&P 500 and the EAFE Index, a widely cited international equity index.

On the surface historical hedge fund performance has been good, but there are a number of caveats to keep in mind. Most hedge fund databases are rife with a number of biases. Perhaps the two most important are survival bias and backfill bias. Survival bias refers to the likelihood that mostly funds with good performance will report their returns to the database vendors and underperforming funds will decline to report their numbers. This survival bias has the effect of skewing hedge fund databases with winning funds and under-populating them with losing ones. Backfill bias refers to the case where new and successful funds decide to report only after they have developed a successful track record. Of course, new and unsuccessful funds will likely never report to the hedge fund database vendors. The net result is that some hedge fund databases are "updated" with the strong historical results, thereby resulting in another upward bias to the reported hedge fund returns.

Park et al. (1999) estimate that the various hedge fund biases trim reported hedge fund index returns by approximately 2–3% per annum. There are a number of

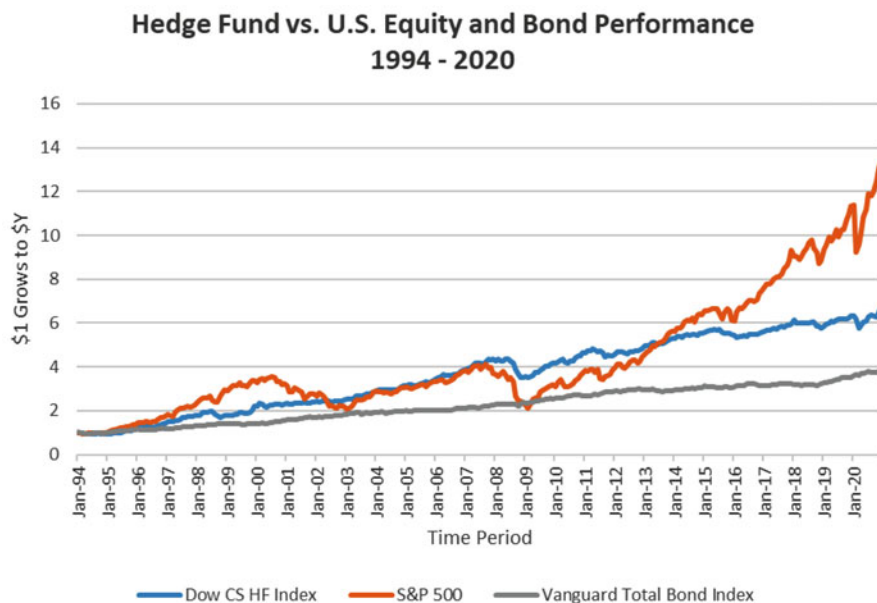


Fig. 1 Hedge fund versus the US equity and bond performance, 1994–2020. Sources: Dow Jones Credit Suisse, Standard and Poors, and Vanguard

popular hedge fund databases, including those managed by Pertrac, Hedge Fund Research, Barclays, Dow Jones Credit Suisse, and Morningstar. Prices for hedge fund databases typically range from \$1,500 to \$20,000 per year.

Notwithstanding the biases noted above, hedge funds in aggregate have demonstrated good historical performance in both nominal and risk adjusted returns, relative to the US equities and bond portfolios. When discussing hedge fund performance, it should be noted that hedge funds are far from a homogeneous asset class. Ideally, a more granular or “apples to apples” comparison should be conducted in order to properly assess the performance of an individual hedge fund. Figure 1 illustrates the performance of the Dow Jones Credit Suisse Hedge Fund Index, from its inception in January 1994 through December 2020, as well as for the S&P 500 and Vanguard Total Bond Market over the same period. Hedge funds had annualized returns of 7.4% over this time period relative to the 10.1% return for the S&P 500 Index, and the 5.1% return for the Vanguard Bond Index Fund.

Additionally, the Dow Credit Suisse Hedge Fund Index had a lower annualized standard deviation (6.8%) than the S&P 500 (12.8%), although a higher value than the Vanguard Bond Index (3.6%). The biggest monthly drawdowns for the Dow Jones Credit Suisse Hedge Fund Index, S&P 500, and Vanguard Bond Index during the aforementioned time period were 7.6%, 20.2%, and 3.3%, respectively. Importantly the correlation between the Dow Credit Suisse Hedge Fund Index and S&P 500 was 0.29. Similarly, the correlations between the Dow Credit Suisse Hedge Fund Index and the Vanguard Bond Index was 0.15. In aggregate, these results

Table 2 Hedge fund versus equity and bond summary statistics, 1994–2020

	Dow CS	S&P	Vanguard
	HF Index	500	Bond Index
Annual return	7.3%	10.1%	5.0%
Annual StdDev	6.8%	12.8%	3.6%
Return/StdDev	107.4%	78.9%	140.4%
Max monthly drop	−7.6%	−20.2%	−3.3%

suggest the Markowitz Efficient Frontier of a diversified portfolio that includes hedge funds may result in superior optimal portfolios. Summary risk and return statistics are shown in Table 2 below.

Correlation of monthly returns, 1994–2020

	Dow CS HF Index	S&P 500	Vanguard Bond Index
Dow CS HF Index	1		
S&P 500	0.286145029	1	
Vanguard Bond Index	0.146184177	0.106580141	1

Sources: Dow Jones Credit Suisse, Standard and Poors, and Vanguard

54.6 Common Types of Hedge Funds

The hedge fund industry has grown substantially since the early days of A.W. Jones. There are approximately 18,000 hedge funds that manage \$3.8 trillion in assets at the end of 2020, according to Preqin.⁹ Hedge Fund Research, Inc. (HFRI) tracks the performance of more than 100 hedge fund categories, including the 29 shown in Table 3. Clearly there is some overlap among the strategies, but the richness of diversity in the universe indicates that it would be a mistake to refer to hedge funds as a homogenous category.

Rather than describe every single hedge fund strategy, our goal here is to discuss some of the most important ones, such as Long/Short and Global Macro. Some investors access hedge funds through a fund of funds vehicle, likely because of the “instant diversification” nature of this investment structure. Of course, the chief drawback is the second layer of fees that must be paid by investing through the fund of funds vehicle. Below we list a brief description of some of the most popular hedge fund strategies.

Activist funds typically purchase a sizeable (i.e., 3% or more) stake in a company and attempt to unlock shareholder value by spurring management or corporate strategy changes, if not the outright sale of the company. In many respects, they operate similar to private equity firms, with the obvious exception that they own a limited number of shares in publicly traded companies. Carl Icahn of Icahn

⁹Preqin is the source of the number of hedge funds and total assets under management for the industry.

Table 3 Common hedge fund categories and corresponding indexes

HFRI Fund Weighted Composite Index	HFRI RV: Fixed Income-Convertible Arbitrage Index
HFRI Equity Hedge (Total) Index	HFRI RV: Fixed Income-Corporate Index
HFRI EH: Equity Market Neutral Index	HFRI RV: Multi-Strategy Index
HFRI EH: Quantitative Directional	HFRI RV: Yield Alternatives Index
HFRI EH: Sector - Energy/Basic Materials Index	HFRI Fund of Funds Composite Index
HFRI EH: Sector - Technology/Healthcare Index	HFRI FOF: Conservative Index
HFRI EH: Short Bias Index	HFRI FOF: Diversified Index
HFRI Event-Driven (Total) Index	HFRI FOF: Market Defensive Index
HFRI ED: Distressed/Restructuring Index	HFRI FOF: Strategic Index
HFRI ED: Merger Arbitrage Index	HFRI Emerging Markets (Total) Index
HFRI ED: Private Issue/Regulation D Index	HFRI Emerging Markets: Asia ex-Japan Index
HFRI Macro (Total) Index	HFRI Emerging Markets: Global Index
HFRI Macro: Systematic Diversified Index	HFRI Emerging Markets: Latin America Index
HFRI Relative Value (Total) Index	HFRI Emerging Markets: Russia/Eastern Europe Index
HFRI RV: Fixed Income-Asset Backed	

Source: Hedge Fund Research, Inc. (www.hfr.com)

Enterprises is one of the more notable activist investors, who were pejoratively termed “corporate raiders” a few decades ago. Bill Ackman of Pershing Square, Dan Loeb of Third Point, and Paul Singer of Elliot Management are other prominent activist investors.

Capital structure arbitrage is a more generalized version of the convertible bond arbitrage trade, discussed below, and is typically executed through the use of complex derivative instruments, such as a credit default swap. As with most arbitrage trades, either side of the position can be taken, but the typical transaction combination is to be long the bond (through a synthetic position in a credit default swap) and short the common stock. The strategy performs well in most circumstances but has difficulty when there is a divergence between stock and bond prices.

Convertible arbitrage funds typically buy convertible bonds and sell short the common stock of the same company. The hedge fund earns the coupon from the bond and the proceeds from the short sales of the common stock, resulting in a “double carry.” The trade is not risk free since dividends may be owed from shorting the common stock and significant losses can be incurred if the bond defaults and the hedge ratio was incorrectly calculated.

Distressed/credit funds focus on companies that are at risk of default, are in default, or have recently emerged from default. Credit is a more general term for the investment in fixed income securities. For example, credit funds may purchase a pool of loans from an investment bank, such as Citi or Goldman Sachs, at deep discounts. Both distressed and credit strategies place a heavy emphasis on legal

work, financial statement analysis, and the identification of the “fulcrum” security, or the one that has the maximum voting leverage in the event of a financial reorganization.

Event driven funds engage in trades based on a specific happening such as a merger, special dividend payment, earnings announcement, analyst opinion upgrade, or credit downgrade. Once the event is resolved, favorably or unfavorably, the fund usually moves on to the next trade.

Fixed income arbitrage funds purchase fixed income securities that appear to be undervalued and sell short other fixed income securities, with similar risk, that appear to be overvalued. Leverage is nearly always used to increase both the risk and return of the trade. A common example would be to purchase a basket of government sponsored mortgage backed securities and sell short a portfolio of Treasury securities with a similar duration. The fund captures the spread (since Treasuries almost always sell at a premium price or lower yield to mortgage backed securities) times the leverage factor, minus the cost of borrowed funds. The trade may backfire during times of market distress, as in the period of The Great Recession.

Global Macro funds do not primarily focus on individual securities but rather emphasize positions in derivatives and index securities on a global basis. They are often unhedged and utilize leverage, making them, in theory, one of the higher risk hedge fund categories.

Long/Short funds purchase securities perceived to be undervalued while simultaneously selling short those viewed as being overvalued. The long to short, or hedge ratio, variation is based on the manager’s outlook. The strategy is most commonly employed for equity related securities but is flexible enough to apply to nearly any asset class.

Managed futures funds trade mainly in futures contracts, such as those in commodities, currencies, metals, materials, or index futures. Some analysts call these funds “Commodity Trading Advisors” (CTAs), but the “managed futures” term better signifies the broad array of instruments that these funds may employ. The name “managed” refers to separately managed accounts, in which fund managers replicate their investment strategies in an account owned and viewed in real time by the client. This structure is in contrast to the limited partnership format of most hedge funds, which provides less transparency and exposes clients to a higher degree of fraud risk since the money is transferred out of their name and into an account controlled by the hedge fund manager.

Market neutral funds balance the market risk of both the long and short side of the portfolio. A common market neutral trade is a long position in Ford and short position in General Motors. The term typically applies to equity-oriented funds, but it could also apply to fixed income securities where the duration (i.e., effective maturity) of the fund’s long and short positions are balanced.

Merger arbitrage funds typically purchase a basket of stocks that are the target of a merger or acquisition. The offsetting trade, in a stock for stock transaction, usually involves a short position in the bidding firm. The main risk of the strategy is the loss incurred when a deal falls through, effectively forcing the hedge fund manager to

absorb the loss of the takeover premium. Leverage is often utilized to increase both the risk and return of the strategy.

Short biased funds are either entirely short or have the bulk of their assets sold short. Since the equity and fixed income markets generally trend up over time, few hedge fund managers successfully operate in this category over the long term. Those that are successful typically focus on small to mid-cap firms with company specific problems.

Statistical arbitrage funds typically employ high frequency computerized trading techniques, or other longer dated quantitative strategies, in order to profit from apparent arbitrage opportunities. True arbitrage opportunities, such as buying IBM in New York and simultaneously shorting it in London for a profit, have virtually no risk and are rare, fleeting, and small in magnitude, if they do exist. The term *statistical* refers to the fact that there may be an arbitrage opportunity according to the fund manager's model; but if the model is not accurate or exogenous factors, such as a takeover, occur, there remains risk in the trade.

Multistrategy funds simultaneously engage in more than one of the previously discussed, or other, hedge fund strategies. They attempt to diversify risk by holding a range of strategies and security positions, rather than placing all of their capital in a single approach. For example, a merger arbitrage fund would have difficulty attracting capital in the event of a slow period for deals. Conversely, a multistrategy fund could allocate capital towards those hedge fund categories that appear to have the best chance of prospering in the current market environment.

Fund of funds purchase positions in a number of individual hedge funds. They attempt to reduce the damage to the portfolio through diversification and rigorous manager selection, in the event that a specific fund "blows up," such as with Long-Term Capital Management. One material disadvantage is that their investors incur a second layer of fees, typically 1% of assets under management and 10% of profits.

54.7 Organization and Regulation of Hedge Funds

Hedge fund organization and regulation differs by country, so we will focus on hedge fund organization and regulation in the USA since it is by far the largest hedge fund market. The Investment Company Act of 1940 provides for certain exemptions that allow hedge funds to avoid much regulation. There is a catch – hedge funds can primarily be offered to a limited number of affluent individual or institutional investors and they cannot advertise in public, except under limited circumstances that we will briefly discuss. The regulatory logic is that if a *small* group of rich individuals or institutions are hurt, it should not be a major blow to the health of the financial system. However, as seen with the Long-Term Capital Management, Madoff, and Archegos debacles, this viewpoint may be flawed. In addition, the regulators believe that high net worth individual and institutional investors, and their advisors, often have the resources to conduct extensive due diligence prior to their investment in hedge funds.

Early hedge funds were organized under the 3(c)(1) exemption of the Investment Company Act of 1940. The 3(c)(1) funds cannot have more than 100 investors and are typically organized under a limited partnership (LP) or limited liability company (LLC) structure. In the USA, it is problematic to charge incentive fees to investors unless they are “accredited,” although the Jumpstart Our Business Startups Act (JOBS Act) of 2012, discussed further below, has changed this definition somewhat. Accredited investors, under Rule 215 of the Securities Act of 1933 and Rule 501 of Regulation D, must have a net worth in excess of \$1 million, or income greater than \$200,000 for the past 2 years, with the expectation of doing so in the future. The income figure is raised to \$300,000 for couples. The Dodd Frank Act of 2010 explicitly excludes the investor’s primary residence in the \$1,000,000 net worth figure for accredited investors.

Many hedge funds have two incarnations of their fund, an onshore fund and offshore fund. This distinction is mainly due to tax reasons, rather than a fundamental difference in investment strategy. Onshore funds are designed to serve domestic, taxable investors. Offshore funds, frequently domiciled in tax haven countries, serve the needs of foreign investors or tax-exempt domestic investors (e.g., pension funds). To a certain extent, funds could circumvent the 100-person limit by creating new funds with slightly different investing objectives. For example, a hedge fund management firm running a Long/Short domestic equity strategy could create a second fund that is focused on Long/Short equities in a *global* investment universe. Furthermore, many hedge funds employ a “master feeder” structure that enables partnerships that consists of many investors to occupy a single slot in the 3(c)(1) fund. Even with these “workarounds,” funds lobbied the SEC for a structure that would enable them to serve more investors.

Section 3(c) 7 of the Investment Company Act of 1996 permits a fund to have up to 500 investors, provided that they are “qualified.” Qualified individual investors own at least \$5 million in investments, excluding home equity. Qualified institutional investors generally have at least \$25 million in net assets. Standard hedge fund terms include a 1-year “lockup” provision, asset based fees of 1–2% per year and incentive fees of 20% per year. However, since hedge funds are private agreements, these terms may vary significantly. For example, it is not uncommon for some of the most successful managers to have a lockup provision of 3–5 years and an incentive fee in excess of 30% per year.

Most hedge funds also include a “highwater mark” provision. This provision prevents hedge fund managers from receiving incentive fees if they have previously lost money, until the client is “made whole” with subsequent positive returns. Asset management fees are charged regardless of investment performance. The incentive fees are typically taken annually (e.g., on December 31), while the asset management fees are usually taken near the end of each calendar quarter and are often used to fund operating expenses.

Investment managers, including hedge funds, which control more than \$100 million in assets, are required to file Form 13F on a quarterly basis with the SEC. Form 13F lists the long equity and option holdings of a fund. Interestingly, short sale positions are currently not required to be disclosed, as well as more “exotic”

holdings of a fund, such as credit default swaps. As we shall discuss later, this “piecemeal” disclosure policy is likely to change. The Dodd Frank Act of 2010 (“Dodd Frank”) has brought hedge funds under the “too big to fail” umbrella, so large funds will now be under the purveyance of The Federal Reserve. The SEC proposed Regulation Private Fund (“Reg PF”) to be part of the implementation of Dodd Frank. Reg PF was in fact incorporated into the final Dodd Frank Act and requires complete the position disclosure of large hedge funds, with the information available only to the government on a confidential basis. In addition, nearly all hedge funds will be required to register as investment advisors, making them subject to additional paperwork and inspections. Generally, funds managing \$25 million in the USA will have to register with the state of their principal office. Once the fund passes the \$100 million threshold, it will be required to register at the federal level with the SEC.

The Jumpstart Our Business Startups Act (JOBS Act) of 2012, which was briefly noted earlier, included a number of provisions that were friendly to the hedge fund industry. Under certain provisions, hedge funds may now accept up to 2,000 investors, including up to 500 that are “unaccredited.” The Act also enabled hedge funds to advertise and maintain a social media presence, containing both public and private sections. Specific information about the hedge fund must be in the password protection portion of the site and involves a prospective client screening process in order to obtain the password. Hedge funds must also take care with how they present performance and are encouraged to show returns net of all fees, since the difference between gross and net returns may be quite substantial.

54.8 Sample Hedge Fund Strategies

There are virtually an unlimited number of hedge fund strategies ranging from simple long only equity or fixed income holdings to exotic, custom tailored derivative trades. Our focus here is to discuss a few strategies that are common in the hedge fund world, yet also provide insight into how hedge funds view risk and return. We will work through three sample trades, or brief case studies.

54.8.1 Case 1: Dual Share Class Arbitrage

Arbitrage is a common hedge fund strategy, in its multitude of forms. All arbitrage trades, in essence, identify two securities or portfolios with similar risk characteristics, but with different expected returns. Arbitrage trades involve purchasing the security or portfolio with the high expected return and shorting the corresponding investment(s) with the low expected return. Leverage is often used to increase the risk and return of the trade. A “carry trade” may also be utilized to further increase the risk and expected return of the trade. A carry trade involves borrowing money where it is cheap (e.g., Japan) and investing it in currencies or investments with higher yields.

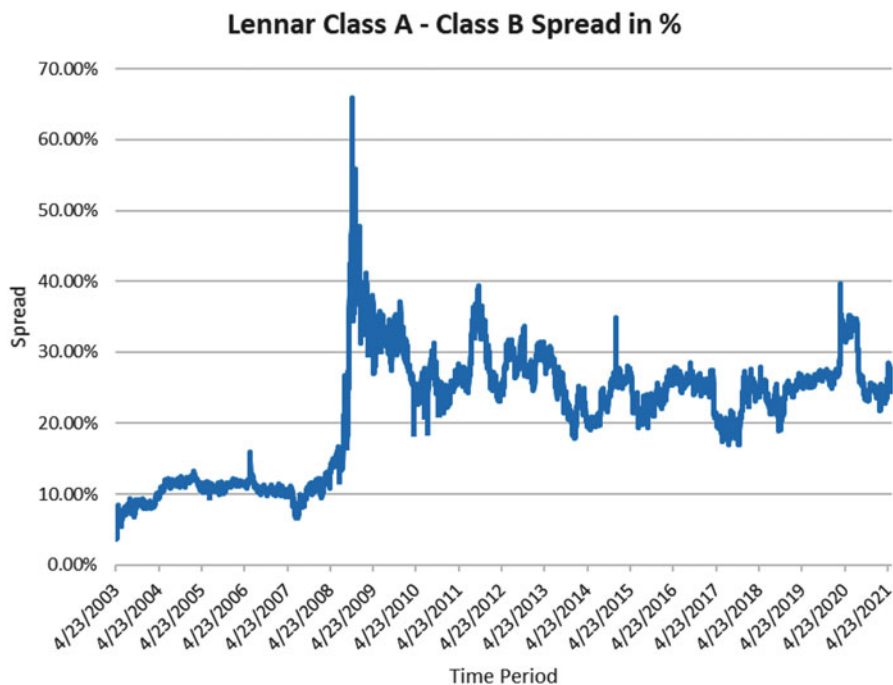


Fig. 2 Dual share class price differences for Lennar Corporation

A relatively “pure” type of arbitrage trade can be found in the analysis of equities with dual shares classes. In this instance, these securities represent two share classes of the same firm that differ in their voting rights, but provide identical claims to earnings and dividends. Dozens of these security pairs exist on American stock exchanges, such as with Alphabet (GOOG, GOOGL). A combination of liquidity and voting rights helps determine the structural relationship between the two share classes.

Usually the difference in prices between the two share classes is small, but occasionally the divergence is quite wide and volatile, possibly creating an arbitrage opportunity. For example, the American homebuilding company, Lennar Corporation, has dual share classes for its equity and its spread has varied dramatically over time. Class B shares have eight times the voting rights of Class A shares, but equal cash flow rights. Figure 2 illustrates that Class B shares consistently sell at a *discount* to their Class A counterparts, despite their superior voting rights and identical cash flow characteristics. Over the past 18 years, the discount has averaged nearly 22% and ballooned further over the course of the 2007–2009 Great Recession. The mystery of the large price differentials is cleared up when the trading volume of each share class is analyzed. Class A shares of Lennar have an average trading volume of approximately 2.5 million shares per day, while Class B

shares average only 59,000 shares per day. Large institutions are, for most practical purposes, limited to trading in the more liquid Class A shares of Lennar.

When the spread exceeds its historical norms, a hedge fund manager may put on a trade going long the B share class of Lennar while shorting its A share class. If the spread approaches its historical value of roughly 22%, then the trade may earn attractive returns for the hedge fund manager. Of course, the trade could go against the manager, as occurred in the fall of 2008, if the spread were to widen. Furthermore, given the sizeable long-term spread between the two share classes, the possibility exists that an activist investor would force the firm to consolidate the two share classes into one, thereby eliminating the spread completely.

54.8.2 Case 2: Pairs Trading with Ford and General Motors

Pairs trading is a form of statistical arbitrage that attempts to identify mispricing between two securities. The term “statistical” refers to the likelihood that there is an apparent arbitrage opportunity, but it is not guaranteed. The pairs, sometimes called twins, are usually equities of firms in the same industry, with similar market capitalization, liquidity, and other fundamental attributes. Mathematically, there should also be a relatively high correlation between the historical price movements of the two companies. In our example, we will focus on Ford (F) and General Motors (GM). Both firms are large capitalization, American based automobile manufacturing companies that are operating on a global scale.

The trade often begins by computing the historical ratio of split adjusted prices between two comparable firms. Some analysts prefer to track the historical ratio of a valuation metric, such as the Price to Earnings or Enterprise Value to EBITDA ratio, in place of the simple price ratio. It is common to utilize at least 1 year of daily data or 5 years of monthly data to estimate the historical relationship. In the analysis below, we utilized 5 years of daily historical data.

In our Ford and GM example, GM’s price historically averaged 3.65 times the daily, split adjusted price of Ford. The standard deviation of this price ratio was 0.55. Assuming a normal distribution, a 95% confidence interval would entail a high or upper bound ratio of 4.76 ($= 3.65 + 2 * 0.55$) and a lower bound ratio of 2.55 ($= 3.65 - 2 * 0.55$). So, roughly 95% of the time there would be *no trade*, since the price ratio would fall within this boundary. If the ratio of the prices is greater than 4.75, then the hedge fund manager would go long Ford and short GM. Conversely, if the ratio of the prices were to fall below 2.55, then the manager would go long GM and short Ford. The manager would close out the trade if the ratio reverted towards its mean of 3.65, or perhaps within one standard deviation of it. If the spread widened, the hedge fund manager would lose money on the trade. Figure 3 shows the price ratio between Ford and GM over the June 2020–June 2021 time period and the aforementioned upper and lower bounds.

Looking at Fig. 3 we can see that a trade (Long Ford, Short GM) would have occurred in early November 2020, since it surpassed the 4.76 upper barrier. Initially the trade would be going against the hedge fund manager, since the ratio continued

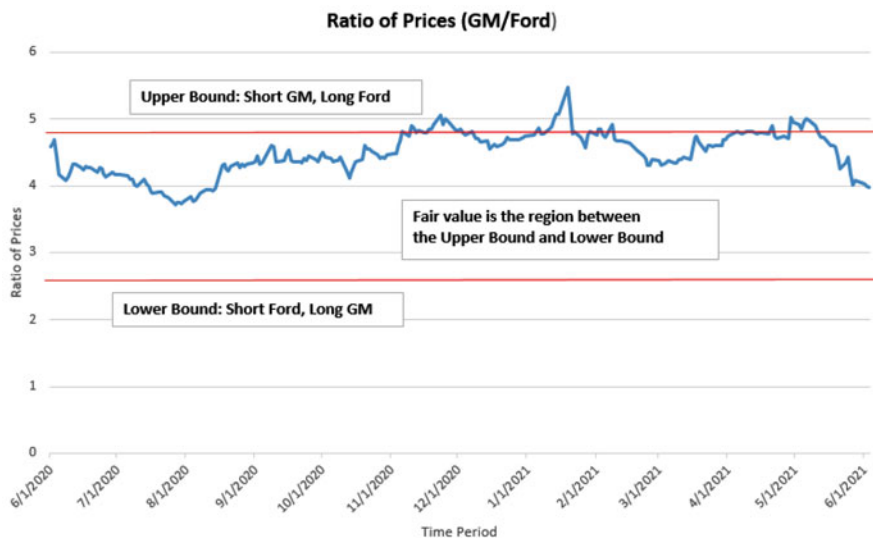


Fig. 3 Graph of historical price ratio of Ford and General Motors

to increase, peaking at roughly 5.5 in mid-January 2021. However, by June of 2021, the ratio fell to approximately 4, resulting in a profit on the trade as the spread shrunk. Presumably, the manager would have closed out the trade somewhere in this time frame or they could continue to wait until it approached the historical mean of 3.65 or even as far as the lower bound value of 2.55.

Pairs trading relies heavily on the concept of reversion to the mean. If the fundamentals of either company have materially changed, then the historical relationship will likely be of little value in predicting future returns. For example, the fundamentals of Ford and GM fundamentals vastly diverged during the Great Recession, with GM eventually filing for bankruptcy, while Ford continued to gain market share and have its share price rise.

54.8.3 Case 3: Activist Investor Carl Icahn's Forced Breakup of Motorola

The first two cases involved the common arbitrage technique of going long one security and short another. Some hedge funds are primarily long only, including the “activist” hedge funds run by legendary investors Carl Icahn, Bill Ackman, and Dan Loeb. During our earlier discussion of activist hedge funds, we noted they often take a material position in a company and try to spur meaningful changes, such as a strategy change, merger, spinoff, or special dividend. Activist investors aim to become a large shareholder in their target firm in order to get the ear of management and often to obtain a seat on the board to spur changes.

Carl Icahn first began buying shares of Motorola for his hedge fund in 2007, acquiring a 4% stake. Investors, or coordinated groups of investors, that reach a 5% ownership threshold must declare their intentions and with the SEC through Form 13D and promptly disclose any changes in ownership levels thereafter. Around 2007, Motorola lagged the performance of other cellphone firms; first, Nokia and Research in Motion/Blackberry, and eventually Apple. Icahn and two of his allies obtained board seats at Motorola to drive change. They then pushed the company to split into two parts, Motorola Solutions and Motorola Mobility. The latter firm was rich in intellectual property and was sold to Google (now Alphabet) for \$12.5 billion in 2011. In 2012, Icahn sold the bulk of his shares in Motorola Solutions back to the firm for \$1.2 billion and resigned from his board seat. All told, Icahn earned a nearly 50% return on his investment in Motorola.¹⁰ In contrast, the S&P 500 returned approximately 15% in aggregate, over the 2007–2012 period, which of course included the devastating bear market of late 2007 to early 2009.

54.9 Trends in Hedge Funds

The hedge fund industry is continuing to evolve. Despite its origins in the early part of the twentieth century, the industry started to “come of age” in the 1990s¹¹. Accordingly, it is still young, relative to many other financial service industries. The turbulent financial markets over the past two decades have accelerated the rate of change in the industry. Listed below are six trends that will, in my view, affect the hedge fund industry.

54.9.1 Trend 1: Further Compression of Fees

The standard hedge fund fee structure consists of a 1–2% asset-based fee and a 20% profit participation or incentive fee. Many funds charge much higher rates and hedge fund of funds charge an additional layer of fees. For example, the financial press routinely reports that Renaissance’s Medallion Fund charges a fee equal to 5% of assets and 44% of profits. The typical fund of fund charges 1% of assets and 10% of profits.

Table 4 shows the gross return, typical fees, and net return for a hypothetical fund of fund. For example, a gross return of 10% shrinks to a *net* return of only 4.5% for the investor – pretax! Hedge fund investors were glad to pay fees when returns were high, or even just modestly positive, as in the devastating 2000–2002 bear market. However, in a muddled market environment, these fees will simply generate too great a drag on, and become too large a percentage of, the total fund returns. In contrast, many index mutual funds and exchange traded funds (ETFs) charge

¹⁰ Kosman (2012) discussed Icahn’s returns on his Motorola investments.

¹¹ Portions of this section are adapted from Eichen and Longo (2009).

Table 4 Gross and net returns for a hypothetical fund of fund

Traditional Fund of Funds		
<i>Hypothetical Gross Return</i>		10.00%
<i>Less:</i>		
Hedge Fund – Base Fee	(1.00%)	9.00%
Hedge Fund – Expenses	(0.65%)	8.35%
Hedge Fund – Incentive Fee (20%)	(1.67%)	6.68%
Fund of Fund – Base Fee	(1.00%)	5.68%
Fund of Fund – Expenses	(0.65%)	5.03%
Fund of Fund – Incentive Fee (10%)	(0.50%)	
Net return to investor		4.53%

asset-based fees of 0.2% or less per annum. Actively managed mutual funds charge asset-based fees usually in the range of 0.5–1.0%.

In addition, given the lack of barriers to entry in the hedge fund space and well-documented challenge in beating low-cost index fund strategies, it is inevitable that the average hedge fund fee will continue to come down. Since institutional investors, who have more buying leverage and are not accustomed to paying “rack” retail rates, are contributing to hedge funds, it is hard to imagine that they will not demand fee concessions from many hedge funds in return for large capital commitments. While established “star” managers may be immune from this fee pressure, it is all but certain that emerging managers will have to succumb to fee compression, thereby creating a trend toward a new and lower fee structure which those who follow will have to comply with. This was the trend in the “long only” institutional money management business, which saw fees decline precipitously over the past 30 years due to increased competition and consolidation. This was also the trend in the mutual fund business where high loads and high internal fees were the norm 20 years ago. The hedge fund business will follow this trend as it continues to evolve. According to Whyte (2021), a recent report by Hedge Fund Research (HFR) found that the average hedge fund fee at the end of 2020 was 1.37% of assets and 16.36% of profits.¹² In my view, “standard” hedge fund fees will approach 1% / 10% within the next ten years. In addition, we are likely to see a focus of payment on alpha not beta return, as discussed in Trend 2.

54.9.2 Trend 2: Payment for Alpha, Not Beta

Trend 1 predicted hedge fund fees will continue to trend lower over time. A nuanced look at the fees paid to hedge funds points to another likelihood. That is, the payment for alpha and not beta. Beta exposure, cheaply and easily achieved through an index fund or ETF, costs close to nothing. For example, the SPDR S&P 500 ETF Trust (SPY) currently has about \$400 billion in assets and charges an annual net expense

¹²Hedge Fund Research provides an estimate of the average hedge fund asset based and incentive fees.

ratio of only 9 basis points (0.09%). Alpha on the other hand, by definition, takes skill and is harder to achieve. Talented hedge fund managers may attempt to differentiate themselves by being paid primarily on alpha and not beta. For example, hedge fund managers may choose to take 50% of the alpha, assuming returns are greater than 0. When legendary investor, Warren Buffett, ran a hedge fund in the 1950s–1960s, he charged no asset-based fee and took 25% of the profits for returns in excess of 6%. Similar fee structures can be devised so the manager is primarily compensated for the hard to deliver alpha versus the relatively costless beta for any particular market.

54.9.3 Trend 3: Increased Liquidity

Some private investment funds, such as those making venture capital investments in developmental stage firms, have a clear and valid need for long-term lockup arrangements. These funds have returns that typically exhibit a “J Curve” effect where returns are negative in early years as capital is put to work, but strongly positive in later years as the investments are harvested. However, most hedge funds have a 1-year lockup provision and advance notice of ninety days prior to any capital withdrawal. Yet, the underlying instruments in most hedge funds are sufficient for 30 days or less liquidity. Other than the desire to earn fees for a longer period of time, it is hard for the average hedge fund to argue that it is absolutely necessary to hold client funds beyond 30 days.

All else being equal, from the investor’s perspective, one does not want to be a year away from accessing their money. The investor wants to be able to withdraw their money in the event that they see something with the fund they do not like, or simply if they need the liquidity. Since most hedge funds do not “need” to have these lockup provisions, as the industry matures, this will be one easy way for managers to distinguish themselves and accommodate client needs. For example, Société Générale’s Lyxor alternative investment platform provides access to hedge funds with attractive liquidity terms and factor level transparency. Thus, we may see increased liquidity becoming the norm of hedge fund investing, not the exception.

54.9.4 Trend 4: Increased Transparency and Growth of Managed Account Hedge Funds

Investing on faith was once accepted as part of the “price of admission” into the hedge fund industry. Indeed, the cachet of an exclusive or closely guarded strategy may have been a positive attribute for some investors. Then came Long-Term Capital Management and the fraud of Bernard Madoff, both of which were discussed earlier. Factor level transparency may not meet the due diligence requirements of many sophisticated retail or institutional investors. Investors have a right to know exactly what they own and how returns are generated. More investors will make these demands on their hedge fund managers in the face of misleading, and in some

cases, fraudulent activity. As such, many funds, especially those newer emerging funds, will be forced to deliver full transparency if they wish to raise substantial assets.

Institutional investors often demand that their hedge fund investments be located in a separately managed account vehicle run *pari passu* with the hedge fund company's traditional partnership structure. To preserve strategy confidentiality, nondisclosure agreements with "teeth" may have to be signed and enforced in return for transparency. As discussed earlier, hedge funds managing more than \$100 million must list their long equity and option holdings on a quarterly basis via Form 13F. Data reported to the SEC, such as through Form 13F, may be of significant value for deciphering the strategy of longer-term equity-oriented hedge funds. However, for active hedge fund managers, the stale data may give some insight into the strategy, but little comfort with respect to a fund's current holdings. In addition, the 13F data do not adequately enable investors to get a good understanding of funds trading in swaps and other complex securities as was noted with the Archegos implosion. Short sellers have come under pressure in light of the dramatic drops in shares of financial companies in 2008. In the opposite direction, several short sellers were badly damaged in 2021 with the dramatic rise of "meme" stocks, such as GameStop and AMC Entertainment.¹³ Hence, greater disclosure may be required for proper due diligence and a safe regulatory environment.

54.9.5 Trend 5: Continued Growth in Hedge Funds, Including Public Vehicles

Hedge funds are likely to continue to experience attractive growth for a number of reasons. First, their risk adjusted returns have been reasonably attractive and the moderate correlation that they exhibit with other asset classes may improve the efficient frontier of client portfolios. Second, the Dodd Frank Act of 2010 limited the amount of proprietary trading that banks can conduct, due to the so-called "Volker Rule." The Rule limits proprietary investments to 3% of the firm's primary – or Tier 1 – capital. For example, Goldman Sachs shuttled its equity proprietary trading unit.¹⁴ Morgan Stanley spun off its Process Driven Trading (PDT) division, historically a leading force behind its equity proprietary trading profits.¹⁵ The natural landing spot for talented prop traders are in hedge funds.

The rise in affluence around the world, most notably in the emerging market "BRIC" countries – Brazil, Russia, India, and China – will result in increased demand for hedge funds. Not only will hedge funds be created in these emerging

¹³Carew (2021) discussed hedge fund losses in GameStop, AMC Entertainment, and other "meme" stocks.

¹⁴CNBC.com discussed Goldman Sachs' closing of its equity proprietary trading unit.

¹⁵Baer (2011) discussed Morgan Stanley's plan to spin off its Process Driven Trading (PDT) division.

markets, but their high net worth investors will demand that hedge funds manage a portion of their substantial wealth. A portfolio of simply “long only” investments can result in a helpless, sinking feeling during bear markets such as in 2008 or the first quarter of 2020. Emerging stock markets, such as those in Russia and China, fell in excess of 50% in 2008, despite not being at the epicenter of the credit products that precipitated The Great Recession. Clearly, affluent investors and institutions in these countries would like to temper this volatility and hedge funds are a natural outlet for some of their capital. Lastly, the lure of high compensation will ensure that hedge funds always remain a preferred destination for the proverbial “best and brightest.”

Although our working definition of a hedge fund focused on private investment vehicles, in recent years a number of hedge fund type products have been offered via public investment vehicles, such as through mutual funds and ETFs. Many of these public hedge fund type vehicles engage in short-selling and derivative transactions, although they still have restrictions concerning leverage, diversification requirements, and liquidity issues. For example, the BlackRock Strategic Income Fund, which currently has about \$40 billion in assets under management, has a global mandate and engages in short-selling and derivative transactions. Asset managers and investors are increasingly looking to include alternative investments in retirement plans, such as IRAs and 401(k)s so these “hedge funds for the masses are likely to gain some traction.”¹⁶

However, not all hedge funds will seek to have public analogs. Public hedge fund type products are currently prohibited from charging incentive fees, so many of the best managers will not run them since their compensation will be materially lower. In addition, some hedge fund managers are uncomfortable with public disclosure of all of their investment positions, despite the likelihood that they will eventually have to do so, as discussed further in the next section.

54.9.6 Trend 6: Increased Regulation

One of the primary objectives of A.W. Jones’ hedge fund was to consciously avoid regulation. The industry has historically escaped significant regulation, but its growth and a number of scandals over the years have made increased regulation inevitable. The Madoff fraud scandal was a particularly embarrassing moment for the SEC, given it had numerous forewarnings regarding Madoff’s shenanigans.¹⁷ However, it should be noted that unlike the “who’s who” of Wall Street firms that needed government bailouts, hedge funds navigated the 2007–2009 Great Recession without government assistance.

¹⁶Bain (2020) discussed the push by alternative investment firms to offer their products in 401(k) and other retirement platforms.

¹⁷Markopolos (2010) brought Madoff’s fraud to the attention of the SEC on more than one occasion, only to have it fall on deaf ears.

The Dodd-Frank Act of 2010 represents an important piece of legislation that resulted in the increased regulation of hedge funds. As noted earlier, all hedge funds of \$100 million or more that are doing business in the USA must now register with the SEC, while funds of \$25 million in size or more have to register with states. Large hedge funds now fall under the “too big too fail” portion of the Act and are subject to oversight by the Federal Reserve. Given the Archegos scandal, it is increasingly likely that in the not too distant future hedge funds will have to disclose *all* investment positions on a quarterly basis in a *public* manner, in a similar manner to mutual funds.

Regulation is also likely to increase on a global basis, reducing the opportunity for regulatory arbitrage. In any event, the new regulations that hedge funds may face will likely be less stringent than those faced by firms in other financial industries, such as Insurance, and will likely not materially impede their continued growth. Hedge funds are unlikely to have their powers significantly curtailed to use leverage, maintain concentrated portfolios, or to hold illiquid assets. Accordingly, they may serve a gap in traditionally diversified portfolios.

54.10 Conclusion

The term hedge fund is hard to precisely define, given the multitude of strategies, risk management processes, and leverage employed by each fund. Nevertheless, hedge funds in their various forms may play a useful role in most individual and institutional portfolio due to the prospect of higher risk adjusted returns, as well as the lack of correlation with traditional stock and bond portfolios. In short, they may improve the Markowitz Efficient Frontier of most portfolios. The industry is still evolving and its continued growth will likely result in the further compression of fees, increased liquidity, improved transparency, wider accessibility, and increased regulation.

References

- Baer, Justin. 2011. Morgan Stanley to spin off prop trading desk. *Financial Times*, January 10. <http://www.ft.com/intl/cms/s/0/6c285214-1cfc-11e0-8c86-00144feab49a.html#axzz1f0sbCI2J>
- Bain, Benjamin. 2020. *Private equity gets a big win with U.S. Nod to Tap 401(k) plans*. Bloomberg. www.bloomberg.com/news/articles/2020-06-03/private-equity-firms-get-trump-administration-s-nod-to-tap-401ks
- Carew, Sinead. 2021. *GameStop, AMC extend Rallies, Gouging short sellers*. Thomson Reuters, May 26. www.reuters.com/technology/gamestop-amc-extend-rallies-gouging-short-sellers-2021-05-26/
- CBS News. 2010. The Madoff Scam: Meet the liquidator. June 20. http://www.cbsnews.com/2100-18560_162-5339719.html?tag=contentMain;contentBody
- Dow Jones Credit Suisse Hedge fund indexes. 2021. <http://www.hedgeindex.com/hedgeindex/en/default.aspx?cy=USD>, Accessed September 29, 2021
- Editor. Alfred Winslow Jones: Hedge Funds, Bio, Quotes, Books. *ValueWalk*, 17 June 2020. www.valuwalk.com/alfred-winslow-jones/

- . Forbes Billionaires 2021: The richest people in the world. *Forbes Billionaires list*, Forbes. www.forbes.com/billionaires/
- Eichen, Mitchell, and John Longo. 2009. Global trends in Hedge funds. In *Hedge fund alpha: A framework for generating and understanding investment performance*, ed. John Longo. Hackensack, NJ: World Scientific Press.
- Grant, David. 2009. David Tepper and Co.: Which traders profited most from the downturn? The Christian Science Monitor, December 21. www.csmonitor.com/Business/2009/1221/David-Tepper-and-Co.-Which-traders-profited-most-from-the-downturn
- Hedge Fund Research, Inc., www.hfr.com, accessed September 29, 2021
- Kosman, Josh. 2012. *Icahn Books \$700M El Paso Payoff*. New York Post, February 28. www.nypost.com/2012/02/28/icahn-books-700m-el-paso-payoff/
- Longo, John. 2009. *Hedge fund alpha: A framework for generating and understanding investment performance*. Hackensack, NJ: World Scientific Press.
- Loomis, Carol. 1966. The Jones nobody keeps up with. *Fortune*, April 1966, pp. 237–247.
- Lowenstein, Roger. 2000. *When genius failed: The rise and fall of long-term capital management*. New York: Random House.
- Mallaby, Sebastian. 2010. *More money than god: Hedge funds and the making of a new elite*. New York: Penguin Press.
- Markopolos, Harry. 2010. *Madoff Whistleblower: SEC failed to do the math*. NPR, March 2. www.npr.org/templates/story/story.php?storyId=124208012
- Norrestad, F. 2021. Assets of pension funds globally 2009–2019. *Statista*, January 21. www.statista.com/statistics/421220/global-pension-funds-assets/
- Park, James, Stephen Brown, and William Goetzmann. 1999. Performance benchmarks and survivorship bias for Hedge fund and commodity trading advisors. *Hedge Fund News*, August. http://www.hedgefundnews.com/news_n_info/article_detail.php?id=108
- Schatzker, Erik, et al. 2021. *Bill Hwang had \$20 billion, then lost it all in two days*. Bloomberg, April 8. www.bloomberg.com/news/features/2021-04-08/how-bill-hwang-of-archegos-capital-lost-20-billion-in-two-days
- Shazar, Jon. 2021. *RenTech can't be wrong, so the world must be*. Dealbreaker, January 14. <https://dealbreaker.com/2021/01/medallion-up-76-percent>
- Statista Research Department. 2021. Global assets of US-based mutual funds 1998–2020. *Statista*, May 18. www.statista.com/statistics/255518/mutual-fund-assets-held-by-investment-companies-in-the-united-states/#:~:text=It%20was%20found%20that%20the,trillion%20U.S.%20dollars%20in%202020
- Whyte, Amy. 2021. *How low will hedge fund fees go in 2021?* Institutional investor, January 4. www.institutionalinvestor.com/article/b1pzd8t3bxdk3x/How-Low-Will-Hedge-Fund-Fees-Go-in-2021
- Zuckerman, Gregory. 2010. *The greatest trade ever: The behind-the-scenes story of how John Paulson defied wall street and made financial history*. New York: Crown Business.



An Appraisal of Modeling Dimensions for Performance Appraisal of Global Mutual Funds

55

G. V. Satya Sekhar

Abstract

A number of studies have been conducted to examine investment performance of mutual funds of the developed capital markets. Grinblatt and Titman (J Bus 62: 393–416, 1989, Am Econ Rev 85:1088–1105, 1994) found that small mutual funds perform better than large ones and that performance is negatively correlated to management fees, but not to fund size or expenses. Hendricks, Patel, and Zeckhauser (Hot hands in mutual funds: a short run persistence of relative performance. J Finance 48:93–130, 1993), Goetzman, Ibbotson, (Do winners repeat; predicting mutual fund performance. J Portfolio Mgmt 20:9–18, 1994), and Brown, Goetzmann (Performance persistence. J Finance, American Finance Association 50(2):679–698, 1995) present evidence of persistence in mutual fund performance. Grinblatt and Titman (The persistence of mutual fund performance. J Finance 47(5):1977–1984, 1992), and Elton et al. (J Fin Econ 42:397–421, 1996) show that past performance is a good predictor of future performance. Blake et al. (The performance of bond mutual funds. J Bus 66:371–403, 1993), Detzler (The performance of global bond mutual funds. J Banking Fin 23: 1195–1217, 1999), and Philpot et al. (Active management, fund size and bond mutual fund returns. Fin Rev 33:115–126, 1998) find that performance is negatively correlated to fund expense, and that past performance does not predict future performance. However, Philpot et al. (Performance persistence and management skill in nonconventional bond mutual funds. Fin Serv Rev 9:247–258, 2000) provide evidence of short-term performance persistence in high-yield bond mutual funds. In their studies of money market mutual funds, Domain and Rechenstein (Performance and persistence in money market mutual fund returns, Fin Serv Rev 6:169–183, 1998) find that the expense ratio is the most important factor in explaining net return differences. Christoffersen (Why do Money fund

G. V. S. Sekhar (✉)

Gandhi Institute of Technology and Management Studies, GITAM, University, Visakhapatnam, AP, India

managers voluntarily waive their fee? J Fin 74:117–140, 2001) shows that fee waivers matter to performance. Smith and Tito (J Fin Quant Anal 4:449–471, 1969) conducted a study into 38 funds for 1958–67 and obtained similar results. Treynor (Harv Bus Rev 43:63–75, 1965) advocated the use of Beta Coefficient instead of the total risk.

Statistical Technique used: This chapter is intended examine the modeling dimensions of measuring performance of mutual funds during the last 50 years, which leads to innovative research in financial modeling of mutual fund's performance measure with the help of various models like: regression model, Treynor Model, Lee Model etc.

1. To understand financial modeling techniques used for performance evaluation of global mutual funds during last 50 years.
2. To develop scope for new model in the area of performance appraisal of mutual funds

Keywords

Financial modeling · Global investments · Mutual funds · Performance appraisal

55.1 Introduction

The measure of performance of financial instruments is basically dependent three important models derived independently by Sharpe, Jensen and Treynor. All three models are based on the assumption that (1) all investors are averse to risk, and are single period expected utility of terminal wealth maximizers, (2) all investors have identical decision horizons and homogeneous expectations regarding investment opportunities, (3) all investors are able to choose among portfolios solely on the basis of expected returns and variance of returns, (4) all trans-actions costs and taxes are zero, and (5) all assets are infinitely divisible.

55.2 Performance Evaluation Methods

The following paragraphs indicate a brief description of the studies on 'performance evaluation of mutual funds'.

Friend et al. (1962)¹ offered the first empirical analysis of mutual funds performance. Sharpe (1964), Treynor and Mazuy (1966), Jensen (1968), Fama (1972), Grinblatt and Titman (1989, 1994) are considered to be classical studies in performance evaluation methods. Sharpe (1964)² made a significant contribution in the

¹Friend et al. (1962).

²Sharpe (1964).

methods of evaluating mutual funds. His measure is based on capital asset prices, market conditions with the help of risk and return probabilities. Sharpe (1966) developed a theoretical measure better known as reward to variability ratio that considers both average return and risk simultaneously in its ambit. It tested efficacy through a sample of 34 open-ended funds considering annual returns and standard deviation of annual return risk surrogate for the period for 1954–1963. The average reward to variability ratio of 34 funds was considerably smaller than Dow Jones portfolio, and considered enough to conclude that average mutual funds performance was distinctly inferior to an investment in Dow Jones Portfolio.³

Treynor (1965)⁴ advocated the use of Beta Coefficient instead of the total risk. He argues that using only naïve diversification, the unsystematic variability of returns of returns of the individual assets in a portfolio typically average out of zero. So he considers measuring a portfolio's return relative to its systematic risk more appropriate.

Teynor and Mazuy (1966)⁵ devised a test of ability of the investment managers to anticipate market movements. The study used the investment performance outcomes of 57 investment managers to find out evidence of market timing abilities and found no statistical evidence that the investment managers of any of the sample funds had successfully outguessed the market. The study exhibited that the investment managers had no ability to outguess the market as a whole but they could identify under priced securities.

Michael C. Jensen (1967)⁶ conducted an empirical study of mutual funds during the period 1954–64 for 115 mutual funds. His results indicate that these funds are not able to predict security prices well enough to outperform a buy-the-market and hold policy. His study ignores the gross management expenses to be free. There was very little evidence that any individual fund was able to do significantly better than which investors expected from mere random chance. Jensen (1968) measured the performance as the return in excess of equilibrium return mandated by Capital Asset Pricing Model. Jensen's measure is based on the theory of the pricing of capital assets by Sharpe (1964) Lintner (1965) and Teynor.

Smith and Tito (1969)⁷ conducted a study into 38 funds for 1958–67 and published results relating to performance of mutual funds. However, Mc Donald (1974) examined 123 mutual funds for 1960–69 measures to be closely correlated more importantly, he found that on an average, mutual funds perform about as well as native 'Buy and Hold' strategy.

Fama (1972)⁸ suggested alternative methods for evaluating investment performance with somewhat finer breakdowns of performance on the stock selection,

³Sharpe (1966).

⁴Treynor Jack (1965).

⁵Treynor Jack and Mazuy (1966).

⁶Jensen (1967).

⁷Born Karn Eric (1983).

⁸Fama Eugene (1972).

market timing, diversification and risk bearing. It devised mechanism for segregation part of an observed investment return due to managers' ability to pick up the best securities at a given level of risk from part that is due to the prediction of general market price movements.

Dunn and Theisen (1983)⁹ study is about ranking by the annual performance of 201 institutional portfolios for the period 1973 through 1982 without controlling for fund risk. They found no evidence that funds performed within the same quartile over the 10-year period. They also found that ranks of individual managers based on 5-year compound returns revealed no consistency.

Eun et al. (1991)¹⁰ reported similar findings. The benchmarks used in their study were the Standard and Poor's 500 Index, the Morgan Stanley Capital International World Index, and a self-constructed index of U.S. multinational firms. For the period 1977–1986, the majority of international funds outperformed the U. S. market. However, they are most failed to outperform the world index. The sample consisted of 19 U. S.-based international funds, and the Sharpe measure was used to assess excess returns.

Barua and Varma (1993)¹¹ have examined the relationship between the NAV and the market price on Mastershares. They conclude that market prices are far more volatile than what can be justified by volatility of NAVs. The prices also show a mean reverting behaviour, thus perhaps providing an opportunity for discovering a trading rule to make abnormal profits in the market. Such a rule would basically imply buying Mastershares whenever the discount from NAV was quite high and selling Mastershares whenever the discount was low.

Droms and Walker (1994)¹² used a cross-sectional/time series regression methodology. Four funds were examined over 20 years (1971–1990), and 30 funds were analyzed for a 6-year period (1985–1990). The funds were compared to the Standard and Poor's 500 Index, the Morgan Stanley Europe, Australia, and Far East Index (EAFE) which proxies non-U. S. stock markets, and the World Index. Applying the Jensen, Sharpe, and Treynor indices of performance, they found that international funds have generally underperformed the U. S. market and the international market. Additionally, their results indicated that portfolio turnover, expense ratios, asset size, load status and fund size are unrelated to fund performance.

Bauman and Miller (1995)¹³ studied the persistence of pension and investment fund performance by type of investment organization and investment style. They employed a quartile ranking technique because they noted that "investors pay particular attention to consultants' and financial periodicals' investment performance rankings of mutual funds and pension funds." They found that portfolios managed by investment advisors showed more consistent performance (measured by quartile

⁹Dunn and Theisen (1983).

¹⁰Eun et al. (1991).

¹¹Barua and Varma (1993).

¹²Droms and Walker (1994).

¹³Bauman and Miller (1995).

rankings) over market cycles and that funds managed by banks and insurance companies showed the least consistency. They suggest that this result may be caused by a higher turnover in the decision-making structure in these less consistent funds. This study controls for the effects of turnover of key decision makers by restricting the sample to those funds with the same manager for the entire period of study.

Volkman and Wohar (1995)¹⁴ extend this analysis to examine factors that impact performance persistence. Their data consists of 322 funds over the period 1980–1989, and shows performance persistence is negatively related to size and negatively related to levels of management fees.

Elton et al. (1996)¹⁵ examined the predictability of stock mutual funds performance based on risk-adjusted future performance. It also demonstrated application of modern portfolio techniques on past data to improve selection, which permitted construction of portfolio funds that significantly outperformed a rule based on the past rank alone. The portfolio so selected was reported to have small, but statistically significant, positive risk-adjusted returns during a period when mutual funds in general had negative risk adjusted returns.

Jayadeve (1996)¹⁶ paper enlightens performance evaluation based on monthly returns. His paper focuses on performance of two growth oriented mutual funds (Mastergain and Magnum Express) on the basis of monthly returns compared to benchmark returns. For this purpose, risk adjusted performance measures suggested by Jensen and Treynor and Sharpe are employed.

Carhart (1997)¹⁷ shows that expenses and common factors in stock returns such as beta, market capitalization, 1-year return momentum, and whether the portfolio is value or growth oriented “almost completely” explain short term persistence in risk-adjusted returns. He concludes that his evidence does not “support the existence of skilled or informed mutual fund portfolio managers.”

Yuxingyan (1999)¹⁸ examined performance of 67 US mutual funds and the S & P 500 index with 10-year daily return data from 1982 to 1992. The S & P index was used as bench mark index. Daily data are transformed into weekly data for computational reasons. In the calculations, it was assumed that the S & P 500 market index is a good one, i.e., it is efficient and its variance is constant.

Arnold et al. (2000)¹⁹ study examines the risk-adjusted returns using Sharpe’s Index, Treynor’s Index and Jensen’s Alpha for five portfolios of international mutual funds during 1985–1994. The bench marks for competition were the U.S. market proxied by the Vanguard Index 500 mutual fund and a portfolio of funds that invest solely in U.S. stocks. The results show that for 1985 through 1994 the portfolio of

¹⁴Volkman and Wohar (1995).

¹⁵Elton et al. (1996).

¹⁶Jayadeve (1996).

¹⁷Carhart (1997).

¹⁸YuxingYan (1999).

¹⁹Redmand et al. (2000).

international mutual funds outperformed the U.S. market and the portfolio of U.S mutual funds.

Rahul Bhargava et al. (2001)²⁰ evaluated the performance of 114 international equity managers over the January 1988 to December, 1997 period. Performance tests are conducted using Sharpe, Jensen performance methodologies. Three major findings are reported. First, international equity managers, on an average, were unable to outperform the MSCI world market proxy during the sample period. Second, geographic asset allocation and equity style allocation decisions enhanced the performance of international managers during the sample period. Third, separately managed funds were outperformed mutual funds.

Sadhak (2003)²¹ study is an attempt to evaluate the performance of Indian mutual funds with the help of data pertaining to: (a) trends in income and expenses, (b) investment yield and risk-associated returns, and (c) returns of Indian mutual funds vis-à-vis returns of other emerging markets.

Bala Ramasamy and Yeung's (2003)²² survey focused on Malaysia where the mutual fund industry started in the 1950s but only gained importance in 1980s with the establishment of government initiated programme. The sample size consisting of 56 financial advisors representing various life insurance and mutual fund companies resulted in 864 different profiles of mutual funds. The cojoint analysis was employed to generate the questionnaire and analyse its results. The results of this survey point to three important factors which dominate the choice of mutual funds. These are consistent past performance, size of funds and costs of transaction.

Chang et al. (2003),²³ identified hedging factor in the equilibrium asset pricing model and use this benchmark to construct a new performance measure. Based on this measure, they are able to evaluate mutual fund managers hedging timing ability in addition to more traditional security selectivity and timing. While security selectivity performance involves forecasts of price movements of selected individual stock, market timing measures the forecasts of next period realizations of the market portfolio. The empirical evidence indicates that the selectivity measure is positive on average and the market timing measure is negative on average.

Alexander (2004)²⁴ has suggested a new dimension called 'modified approach for risk-adjusted performance of mutual funds'. This method can be considered as more powerful, because it allows not only for an identification of active resources, but also for identification of risk. He observed two interesting results: First, it can be shown that in some cases, a superior security selection effect is largely dependent on taking higher risks. Second, even in the small sample analyzed in the study, significant differences appear between each portfolio manager's styles of selection.

²⁰Rahul Bhargava et al. (2001).

²¹Sadhak (2003).

²²BalaRamasamy and Yeung (2003).

²³Jow-Ran Chang et al. (2003).

²⁴Obeid (2004).

Gupta OP and Amitab Gupta (2004)²⁵ published their research on select Indian mutual funds during 4 year period from 1999 to 2003 using weekly returns based on NAVs for 57 funds. They found that fund managers have not outperformed the relevant benchmark during the study period. The funds earned an average return of 0.041 per week against the average market return of 0.035%. The average risk free rate was 0.15% per week indicating that the sample funds have not earned even equivalent to risk-free return during the study period.

Subash Chander and Jaspal Singh (2004)²⁶ considered selected funds during the period Nov 1993 to March, 2003 for the purpose of their study. It was found that Alliance Mutual Fund and Prudential ICICI Mutual funds have posted better performance for the period of study in that order as compared to other funds. Pioneer ITI, however, shown average performance and Tepleton India mutual fund has staged a poor show.

Amit Singh Sisodiya (2004)²⁷ makes comparative analysis of performance of different mutual funds. He explains that, a fund's performance when viewed on the basis of returns alone would not give a true picture about the risk the fund would have taken. Hence, a comparison of risk-adjusted return is the criteria for analysis.

Alberto et al. (2005)²⁸ analyzed the passive role that, implicitly, would place institutional investors in such a context. The study was conducted in Italy using empirical evidence from the Italian stock exchange (Comit Index). This study finds that three factors reduce the freedom of institutional investors to manage their portfolio- the market target size, the fund structure and the bench marking.

Sudhakar and Sasi Kumar (2005)²⁹ made a case study of Franklin Templeton mutual fund. The sample consists of a total 10 growth oriented mutual funds during the period April 2004 to March 2005. NIFTY based on NSE Index was used as the proxy for the market index and each scheme is evaluated with respect to the NSE index to find out whether the schemes were able to beat the market or not. It was found that most of growth oriented mutual funds have been able to deliver better returns than the benchmark indicators. In the sample study, all the funds have positive differential returns indicating better performance and diversification of the portfolio, except two funds with negative differential returns viz., Franklin India Blue Chip Fund, Templeton India Income Fund.

Martin Eling (2006)³⁰ made a remarkable contribution to the theory of 'performance evaluation measures'. In this study, data envelopment analysis (DEA) is presented as an alternative method for hedge fund performance measurement. As an optimization result, DEA determines an efficiency score, which can be interpreted

²⁵ Gupta and Amitabh Gupta (2004)

²⁶ Subash Chander and Japal Singh (2004).

²⁷ Amit Singh Sisodiya (2004).

²⁸ Alberto Bertoni et al. (2005).

²⁹ Sudhakar and Sasi Kumar (2005).

³⁰ Martin Eling (2006).

as a performance measure. An important result of the empirical study is that completely new rankings of hedge funds compared to classic performance measures.

George Comer (2006)³¹ examined the stock market timing ability of two samples of hybrid mutual funds. It was found that the inclusion of bond indices and a bond timing variable in a multifactor Treynor-Mazuy model framework leads to substantially different conclusion concerning the stock market timing performance of these funds relative to the traditional Treynor-Mazuy model find less stock timing ability over the 1981–91 time period provide evidence of significant stock timing ability across the second fund sample during the 1999–2000 period.

Yoon K Choi (2006)³² proposed an incentive compatible portfolio performance evaluation measure. In this model, a risk-averse portfolio manager is delegated to manage a fund, and his portfolio construction (and information-gathering) effort is not directly observable to investors. In which managers are to maximize investors' gross returns net of managerial compensation. He considers the effect of organizational elements such as economics of scale on incentive and thus on performance.

Ramesh Chander (2006),³³ study examined the investment performance of managed portfolios with regard to sustainability of such performance in relation to fund characteristics, parameter stationarity and benchmark consistency. The study under consideration is based on the performance outcome of 80 investment schemes from public as well as private sector for the 5-year period encompassing January 1998 through December 2002. The sample comprised 33.75% of small, 26.75% of medium, 21.25% of large and 18.75% of the giant funds.

Ramesh Chander (2006a)³⁴ study on market timing abilities enables to understand how well the manager has been able to achieve investment targets and how well risk has been controlled in the process. The results reported were unable to generate adequate statistical evidence in support of manager's successful market timing. It persisted across measurement criteria, fund characteristics, and the benchmark indices. However, absence of performance is noted for alternative sub-periods signifying the negation of survivorship bias.

Beckmann et al. (2007)³⁵ found that Italian female professionals do not only assess themselves as more risk averse than their male colleagues, they also prefer a more passive portfolio management compared to the level they are allowed to. Besides, in a competitive tournament scenario near the end of the investment period, female asset managers do not try to become the ultimate top performer when they have outperformed the peer group. However in case of underperformance, the risk of deviating from the benchmark makes female professionals more willing than their male colleagues to seize a chance of catching up.

³¹George Comer (2006).

³²Choi (2006).

³³Ramesh Chander (2006).

³⁴Ramesh Chander (2006a).

³⁵Beckmann et al. (2007).

Gajendra Sidana (2007)³⁶ made an attempt to classify 100 mutual funds employing cluster analysis and using a host of criteria like the 1 year old return, 2 year annualized return, 3 year annualized return, 5 year annualized return, alpha, beta etc. The data is obtained from value-research. The author finds inconsistencies between investment style/objective classification and the return obtained by the fund.

Coates and Hubbard (2007)³⁷ reviewed the structure, performance and dynamics of the mutual fund industry, and showed that they are consistent with competition. It was also found that concentration and barriers to entry are low, actual entry is common and continuous, pricing exhibits no dominant long-term trend, and market shares fluctuate significantly. Their study also focused on 'effects of competition on fee' and 'pricing anomalies'. They suggested legal interventions are necessary in setting fee in mutual funds of United States.

Subha and Bharathi (2007)³⁸ study is carried out for open end mutual fund schemes and 51 schemes are selected by convenient sampling method. NAV's are taken for a period of 1 year from 1st October 2004 to 30th September, 2005. Out of the 51 funds as many as 18 schemes earned higher returns than the market return. The remaining 33 funds however generated lower returns than the market.

Sondhi (2007)³⁹ study analyses the financial performance of 36 diversified equity mutual funds in India, in terms of rates of return, comparison with risk free return, bench mark comparison and risk adjusted returns of diversified equity funds. Fund size, ownership pattern of AMC and type of fund are the main factors considered in this study. The study reveals that private sector is dominating public sector.

Cheng-Ru Wu et al. (2008)⁴⁰ study adopts modified Delphi method and the analytical hierarchy process to design an assessment method for evaluating mutual fund performance. The most important criteria of mutual fund performance should be 'mutual fund style' following is 'market investment environment'. This result indicates investor's focus when they evaluate the mutual fund performance.

Eleni Thanou (2008)⁴¹ study examines the risk adjusted overall performance of 17 Greek Equity mutual funds between years 1997 and 2005. The study evaluated performance of each fund based on the CAPM performance methodology, calculating the Treynor and Sharpe Indexes for the 9 year period as well as for three sub-periods displaying different market characteristics. The results indicated that the majority of the funds under examination followed closely the market, achieved overall satisfactory diversification and some consistently outperformed the market, while the results in market timing are mixed, with most funds displaying negative market timing capabilities.

³⁶Gajendra Sidana (2007).

³⁷Coates John and Hubbard Glenn (2007).

³⁸Subha and Bharati Jaya (2007).

³⁹Sondhi (2007).

⁴⁰Cehng-Ru Wu et al. (2008).

⁴¹Eleni Thanou (2008).

Kajshmi et al. (2008),⁴² studied a sample of schemes in the 8 years period. This study is based on performance evaluation is restricted to the schemes launched in the year 1993 when the industry was thrown open to private sector under the regulated environment by passing the SEBI(Mutual Funds) Regulations 1993. The performance of the sample schemes were in line with that of the market as evident from the positive beta values. All the sample schemes were not well diversified as depicted by the differences in the Jensen alpha and Sharpe's differential return.

Massimo Masa and Lei Zhang (2008)⁴³ found the importance of organizational structure on Asset Management Company of mutual fund. Their study found that more hierarchical structures invest less in firms located close to them and deliver lower performance. An additional layer in hierarchical structure reduces the average performance by 24 basis points per month. At the same time, more hierarchical structures leads to herd more and to hold less concentrated portfolios.

Manuel Ammann and Michael Verhofen (2008)⁴⁴ examined the impact of prior performance on the risk-taking behaviour of mutual fund managers. Their sample taken from US funds starts in Jan 2001 and ends in Dec, 2005. The study found that prior performance in the first half of the year has, in general, a positive impact on the choice of the risk level in the second half of the year. Successful fund managers increase the volatility, the beta, and assign a higher proportion of their portfolio to value stocks, small firms, and momentum stocks in comparison to unsuccessful fund managers.

Onur et al. (2008)⁴⁵ study evaluates the performance of 50 large US-based international equity funds using risk-adjusted returns during 1994–2003. This study provides documentation on the risk-adjusted performance of international mutual funds. The evaluation is based on objective performance measures grounded in modern portfolio theory. Using the methodology developed by Modigliani and Miller in 1997, the study reports the returns that would have accrued to these mutual funds for a 5-year holding period as well as a 10-year holding period. It is evident from the empirical results of this study that the funds with the highest average returns may lose their attractiveness to investors once the degree of risk embedded in the fund has been factored into the analysis.

Qiang Bu and Nelson Lacey (2008)⁴⁶ examined the determinants of US mutual fund terminations and provided estimates of mutual fund hazard functions. Their study found that mutual fund termination correlates with a variety of fund specific variables as well as with market variables such as the S&P 500 index and the short-term interest rate. This was tested with the underlying assumptions of the semi-parametric Cox model and reject proportionality. They also found that different fund

⁴² Kajshmi et al. (2008).

⁴³ Massimo Massa and Lee Xhang (2008).

⁴⁴ Manuel Ammann and Michael Verhofen (2008).

⁴⁵ Onur Arugaslan et al. (2008).

⁴⁶ Qiang Bu and Nelson Lacey (2008).

categories exhibit distinct hazard functions depending on the fund's investment objectives.

Smith (2009)⁴⁷ discussed the size and market concentration of the mutual fund industry, the market entry and exit of mutual funds, the benefits and costs of mutual fund size changes, principal benefits and costs of ownership from fund shareholders' perspective etc. This study is based on data from Morningstar (2009) about US mutual fund industry, which was composed of 607 fund families.

Baker et al. (2009)⁴⁸ investigated the relation between the performance and characteristics of 118 domestic actively managed institutional equity mutual funds. The results showed that the large funds tend to perform better, which suggests the presence of significant economies of scale. The evidence indicates a positive relation between cash holding and performance. They also found evidence in a univariate analysis that expense ratio class is an important determinant of performance, and the results are significant in a multivariate setting using Miller's active alpha as a performance metric.

Khurshid et al. (2009)⁴⁹ studied the structure of the mutual fund industry in India and analyzed the state of competition among all the mutual funds in private sector and public sector. The levels of competition and their trends have been obtained for the periods March 2003 to March, 2009. This study found over-all mutual fund industry is facing a high competitive environment. An increasing trend of competition was observed within Bank-Institution, Private sector foreign and private sector joint venture mutual funds.

Mohit Gupta and Agarwal (2009)⁵⁰ study focused on the portfolio creation and industry concentration of 18 ELSS schemes during April 2006 to April 2007. Mutual fund industry concentration was the variable used in classification or cluster creation. This exercise was repeated each month for the period under study. Finally portfolio performance was compared with index fund, portfolio of three randomly picked funds of the previous month, and the return and risk parameters of ELSS category as a whole.

Amar Ranu and Depali Ranu (2010)⁵¹ critically examined the performance of equity funds and found out the top 10 best performing funds among 256 equity mutual fund schemes in this category. Their considers three criteria for selection: (a) mutual funds having 5 years of historical performance, (b) fund schemes having a minimum of Rs.400 crore of assets under management and (c) fund which have average return more than 22.47. They found that HDFC TOP 200(Growth) option was outperforming among the top 10 best performing equity funds.

⁴⁷ Smith (2009).

⁴⁸ Baker Kent et al. <http://ssrn.com/abstract=1124577>.

⁴⁹ Khurshid et al. (2009).

⁵⁰ Mohit Gupta and Navdeep Aggarwal (2009).

⁵¹ Amar Ranu and Depali Ranu (2010).

Sunil Wahal and Albert Wang (2010)⁵² found impact of the entry of new mutual funds on incumbents using the overlap in their portfolio holdings as a measure of competitive intensity. Their study reveals that funds with high overlap also experience quantity competition through lower investor flows, have lower alphas, and higher attrition rates. These effects only appear after the late 1990s, at which point there appears to be endogenous structural shift in the competitive environment. Their concluding remark is that ‘the mutual fund market has evolved into one that displays the hallmark features of a competitive market’.

55.3 A Review on Various Models for Performance Evaluation

55.3.1 Jensen Model

Given the additional assumption that the capital market is in equilibrium, all three models yield the following expression for the expected one period return on any security (or portfolio) j ⁵³:

$$E(R_j) = RF + \beta_j[E(R_m) - RF] \quad (1)$$

RF = the one-period risk free interest rate.

$\beta_j = cov(j R_j, R_M)/\sigma^2 R_M$ = the measure of risk (hereafter called systematic risk) which the asset pricing model implies is crucial in determining the prices of risky assets.

$E(R_M)$ = the expected one-period return on the “market portfolio” which consists of an investment in each asset in the market in proportion to its fraction of the total value of all assets in the market. It implies that the expected return on any asset is equal to the risk free rate plus a risk premium given by the product of the systematic risk of the asset and the risk premium on the market portfolio.

55.3.2 Fama Model

In Fama’s decomposition performance evaluation measure of portfolio, overall performance can be attributed to selectivity and risk.⁵⁴ The performance due to selectivity is decomposed into net selectivity and diversification. The difference between actual return and risk-free return indicates overall performance:

$$R_p - R_f \quad (2)$$

⁵² Sunil Wahal and Alber (Yan) Wang (2010).

⁵³ Jensen (1967).

⁵⁴ Fama (1972).

Where in

R_p: is actually return on the portfolio, which is monthly average return of fund.

R_f- is monthly average return on treasury bills 91-days.

The overall performance further can be bifurcated into performance due to selectivity and risk

Thus,

$$R_p - R_f = [R_p - R_p(\beta_p) + R_p(\beta_p) - R_f] \quad (3)$$

In other words, Overall performance = selectivity + risk

55.3.3 Treynor and Mazuy Model

Treynor and Mazuy developed a prudent and exclusive model to measure investment managers' market timing abilities.⁵⁵ This formulation is obtained by adding squared extra return in the excess return version of the capital asset pricing model as given below:

$$(R_{pt} - R_{ft}) = \alpha + \beta p(R_{mt} - R_{ft}) + \gamma p(R_{mt} - R_{ft})^2 + e_{pt} \quad (4)$$

Where: R_{pt} – is monthly return on the fund, R_{ft}- is monthly return on 91 days treasury bills, R_{mt}- is monthly return on market index, E_{pt} is error term

This model involves running a regression with excess investment return as dependent variable and the excess market return and squared excess market return as independent variables. The value of coefficient of squared excess return acts as a measure of market timing abilities that has been tested for significance of using *t*-test. Significant and positive values provide evidence in support of the investment manager's successful market timing abilities.

55.3.4 Statman Model

Statman measured mutual funds using the following equation⁵⁶:

eSDAR (excess standard deviation and adjusted return)

$$= R_f + (R_p - R_f)(S_m/S_p) - R_m \quad (5)$$

⁵⁵Treynor et al. (1966).

⁵⁶Statman (2000).

In this formulae: R_f - monthly return on 3-month treasury bills, R_p – monthly return on fund portfolio

R_m - monthly return on the benchmark index, S_p - standard deviation of portfolio p's return and S_m – standard deviation of return on the bench mark index.

This model used for short-term investment analysis. The performance is compared with it bench mark on monthly basis.

55.3.5 Choi Model

Choi provides a theoretical foundation for an alternative portfolio performance measure that is incentive-compatible.⁵⁷ In this model, a risk-averse portfolio manager is delegated to manage a fund, and his portfolio construction (and information-gathering) effort is not directly observable to investors. The fund manager is paid on the basis of the portfolio return that is a function of effort, managerial skill, and organizational factors. In this model, the effect of institutional factors is described by the incentive contractual form and disutility (or cost) function of managerial efforts in fund operations. It focuses on the cost function as an organizational factor (simply, scale factor). It was assumed that the disutility function of each fund is determined by the unique nature of its operation (e.g., fund size) and is an increasing function of managerial effort at an increasing rate.

55.3.6 Elango Model

Elango model is also compares the performance of public sector funds vs private sector mutual funds in India.⁵⁸ In order to examine the trend in performance of NAV during the study period, growth rate in NAV was computed. The growth rate was computed based on the following formula:

$$\text{Growth Rate : } R_g = (Y_t - Y_0/Y_0) \times 100 \quad (6)$$

R_g : Growth rate registered during the current year

Y_1 : yield in current year

Y_0 : yield in previous year

In order to examine whether past is any indicator of future growth in the NAV six regression analyses were carried out. NAV of base year was considered as the dependent variable and current year as in the independent variable.

⁵⁷ Choi (2006).

⁵⁸ Elango (2003).

$$\begin{aligned}
 \text{Equation : } Y &= A + bX \\
 \text{Dependentvariable : } Y &= NAV \text{ of } 1999 - 2000 \\
 \text{Independentvariable : } X &= NAV \text{ of } 2000 - 01
 \end{aligned} \tag{7}$$

In the same way, the second regression equation computed using NAVs of 2000–01 and 2001–02, as dependent and independent variables.

55.3.7 Chang, Hung and Lee Model

The pricing model adopted by Jow-Ran chang, Nao-Wei Hung and Cheng-Few Lee is based on competitive equilibrium version of intemporal asset pricing model derived in Campbell.⁵⁹ The dynamic asset pricing model incorporates hedging risk as well as market. This model uses a loglinear approximation to the budget constraint to substitute out consumption from a standard intertemporal asset pricing model. Therefore, asset risk premia are determined by the covariances of asset returns with the market return and with news about the discounted value of all future market returns. Formally, the pricing restrictions on asset i imported by the conditional version of the model are:

$$E_t r_{i,t+1} - rf_{t+1} = -V_{it}/2 + \gamma V_{im} + (\gamma - 1)V_{ih} \tag{8}$$

Where:

$Etri, t + 1$; log return on asset, $rf, t + 1$ log return on riskless asset, V_{ii} denotes $Vart(ri, t + 1)$, γ is the agent's coefficient of relative risk aversion, V_{im} denotes $Covt(ri, t + 1, rm, t + 1)$, and $V_{ih} = Covt(ri, t + 1, (Et + 1 - Et), -\infty j = 1 \rho jrm, t + 1 + j)$, the parameter: $\rho = 1 - \exp(c - w)$ and $c - w$ is the mean log consumption to wealth ratio.

This states that the expected excess log return in an asset, adjusted for a Jensen's inequality effect, is a weighted average of two covariances: the covariance with the return from the market portfolio and the covariance with news about future returns on invested wealth. The intuition in this equation that assets are priced using their covariances with the return on invested wealth and future returns on invested wealth.

55.3.8 MM Approach

Leah Modigliani and Franco Modigliani better known as M² in the investment literature. This measure is developed adjusting portfolio return.⁶⁰ This adjustment is carried on the uncommitted(cash balances) part of the investment portfolio at the risk-less return so as to enable all portfolio holdings to participate in the return

⁵⁹ Chang et al. (2003).

⁶⁰ Modigliani and Modigliani (1997).

generation process. This adjustment is needed to bring out the level playing field for portfolio risk-return and vis-à-vis market return. The effect of this adjustment is reported below:

$$M^2 = {}^*R_p - R_m \quad (9)$$

$${}^*R_p = (R_f^*(1 - S_{dm}/S_{dp})) + (R_p^*S_{dm}/S_{dp}) \quad (10)$$

Table 1 Overview of different measures

Measures	Description	Interpretation
Sharpe ratio	Sharpe Ratio = Fund return in excess of risk free return/Standard deviation of Fund. Sharpe ratios are ideal for comparing funds that have a mixed asset classes	The higher the Sharpe ratio, the better a funds returns relative to the amount of risk taken
Treynor ratio	Treynor ratio = Fund return in excess of risk free return/Beta of Fund. Treynor ratio indicates relative measure of market risk	The higher the Treynor ratio shows higher returns and lesser market risk of the fund
Jensen measure	This shows relative ratio between alpha and beta	Jensen measure is based on systematic risk. It is also suitable for evaluating a portfolio's performance in combination with other portfolios
M ² measure	It matches the risk of the market portfolio and then calculate appropriate return for that portfolio	A high value indicates that the portfolio has outperformed and vice versa
Jensen model	$E(R_j) = R_f + \beta_j[E(R_m) - R_f]$	The expected one-period return on the "market portfolio" which consists of an investment in each asset in the market in proportion to its fraction of the total value of all assets in the market
Fama model	$R_p - R_f = [R_p - R_p(\beta_p + R_p(\beta_p - R_f))]$	Overall performance = selectivity + risk
Treynor and Mazuy model	$(R_{pt} - R_{ft}) = \alpha + \beta p(R_{mt} - R_{ft}) + \gamma p(R_{mt} - R_{ft})^2 + \epsilon_{pt}$	This model involves running a regression with excess investment return as dependent variable and the excess market return and squared excess market return as independent variables
Statman model	$eSDAR = R_f + (R_p - R_f)(S_m/S_p) - R_m$	This model used for short-term investment analysis. The performance is compared with it bench mark on monthly basis
Elango model	$R_g = (Y_t - Y_0/Y_0) \times 100$	In order to examine whether past is any indicator of future growth in the NAV six regression analyses were carried out. NAV of base year was considered as the dependent variable and current year as in the independent variable

In this formulae: * R_p = expected return, R_f = Risk free return, S_{dm} = Standard deviation of market portfolio and S_{dp} = Standard deviation of managed portfolio

In case the managed portfolio has twice the standard deviation of the market, then, the portfolio would be half invested in the managed portfolio and remaining half be invested at the risk-less rate. Likewise, in case the managed portfolio has lower standard deviation than the market portfolio, it would be levered by borrowing money and investing the money in managed portfolio. Positive M^2 value indicate superior portfolio performance while the negative indicates actively managed portfolio manager's inability to beat the benchmark portfolio performance (Table 1).

55.4 Conclusion

This chapter is intended to examine various performance models derived by financial experts across the globe. A number of studies have been conducted to examine investment performance of mutual funds of the developed capital markets. The measure of performance of financial instruments is basically dependent three important models derived independently by Sharpe, Jensen and Treynor. All three models are based on the assumption that (1) all investors are averse to risk, and are single period expected utility of terminal wealth maximizers, (2) all investors have identical decision horizons and homogeneous expectations regarding investment opportunities, (3) all investors are able to choose among portfolios solely on the basis of expected returns and variance of returns, (4) all trans-actions costs and taxes are zero, and (5) all assets are infinitely divisible.

References

- Amar Ranu, and Depali Ranu (2010), The best performing equity funds. *The Analyst*, Jan.
- Amit Singh Sisodiya. 2004. Mutual Fund Industry in India- Coming of age. *Chartered Financial Analyst*, July, p 17–22.
- Ammann, Manuel, and Michael Verhofen. 2008. The impact of prior performance on the risk-taking of mutual fund Manager. *Annals of Finance* 5: 69–90.
- Arugaslan, Onur, Ed Edwards, and Ajay Samant. 2008. Risk-adjusted performance of International Mutual Funds. *Managerial Finance* 34 (1): 5–22.
- Baker Kent H., John A. Haslem, and David M. Smith. 2009. Performance and characteristics of actively managed institutional equity mutual funds. Electronic copy source: <http://ssrn.com/abstract=1124577>
- Bala, Ramasamy, and Matthew C.H. Yeung. 2003. Evaluating mutual funds in an emerging market: Factors that matter to Financial Advisors. *International Journal of Bank Marketing*, March, pp.122–136.
- Barua, S.K., and J.R. Varma. 1993. "Speculative dynamics: The case of master shares", *Advances in financial planning and forecasting*. Vol. 5. Greenwich: Jai Press.
- Bauman, W.S., and R.E. Miller. 1995. Portfolio performance rankings in stock market cycles. *Financial Analysts Journal* 51: 79–87.
- Beckmann, D., T. Lutje, and L. Rebeggiani 2007. Italian Asset Managers' behavior: Evidence on overconfidence, risk taking, and gender. Discussion paper no. 358, Leibniz Universitat Hannover, Department of Economics, ISSN: 0949–9962.

- Bertoni, Alberto, Giorgio Brtmitti, and Chiara Cesari. 2005. Mutual fund bench marking and market bubbles-a behavioural approach. *Transition Studies Review* 12 (1): 36–43.
- Bhargava, Rahul, Jong Gallo, and Peggy T. Swason. 2001. The performance, asset allocation and investment style of International Equity Manager. *Review of Quantitative Finance and Planning* 17: 377–395.
- Born, Kam Eric. 1983. *International banking in the 19th and 20th centuries*. New York: St. Martin's Press.
- Bu, Qiang, and Nelson Lacey. 2008. On understanding mutual fund terminations. *Journal of Economics and Finance* 33: 80–99.
- Carhart, M.M. 1997. Persistence in mutual fund performance. *Journal of Finance* 52: 57–82.
- Cehng-Ru, Wu, Hsin-Yuan Chang, and Wu Li-Syuan. 2008. A framework of assessable mutual fund performance. *Journal of Modeling in Management* 3 (2): 125–139.
- Ramesh Chander. 2006. Informational efficiency, parameter stationarity and Bench Mark Consistency of investment performance. *The ICFAI Journal of Applied Finance*, March.
- Chander, Ramesh. 2006a. Investment Manager's market timing abilities: Empirical evidences. *The ICFAI Journal of Applied Finance* 12 (8): 15–31.
- Chang, Jow-Ran, Mao-Wei Hung, and Cheng-few Lee. 2003. An intemporal CAPM approach to evaluate mutual fund performance. *Review of Quantitative Finance and Accounting* 20: 425–433. 414–433.
- Choi, Yoon K. 2006. Relative portfolio performance evaluation and incentive structure. *Journal of Business* 79 (2): 903–921.
- Coates John C., and R. Hubbard Glenn. 2007. Competition in the mutual fund industry: Evidence and implications for policy. Discussion paper No.592, Aug, Source: <http://ssm.com/abstract=1005426>
- Comer, George. 2006. Hybrid mutual funds and market timing performance. *Journal of Business* 79 (2): 771–797.
- Droms, W.G., and D.A. Walker. 1994. Investment performance of International Mutual Funds. *Journal of Financial Research* 17 (Spring): 1–14.
- Dunn, P.C., and R.D. Theisen. 1983. How consistently do active managers win? *Journal of Portfolio Management* 9: 47–51.
- Elango R. 2003. Which fund yields more returns? A comparative analysis on the NAV performance of select public v/s private/foreign open-ended mutual fund schemes in India, Mutual Funds.
- Eleni, Thanou. 2008. Mutual fund evaluation during up and down market conditions: The case of Greek Equity Mutual Funds. *International Research Journal of Finance and Economics* (13).
- Eling, Martin. 2006. Performance measurement of Hedge Funds using data envelopment analysis. *Financial Markets and Portfolio Management* 20: 442–471.
- Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake. 1996. Market timing ability and volatility implied in investment Newsletters' Asset Allocation Recommendations. *Journal of Financial Economics* 42: 397–421.
- Eun, C.S., R. Kolodny, and B.G. Resnick. 1991. U.S. based international mutual funds: A performance evaluation. *The Journal of Portfolio Management* 17 (Spring): 88–94.
- Fama, Eugene F. 1972. Components of investment performance. *Journal of Finance* 27: 551–567.
- Friend, I., F.E. Brown, E.S. Herman, and D. Vickers. 1962. *A study of mutual funds*. Washington, DC: U.S. Government Printing Office.
- Grinblatt, and Titman. 1989. Mutual fund performance – An analysis of quarterly holdings. *Journal of Business* 62: 393–416.
- Grinblatt, and Titman. 1994. Momentum Investment strategies – Portfolio performance and herding: A study of mutual fund behavior. *American Economic Review* 85: 1088–1105.
- Gupta, Mohit, and Navdeep Aggarwal. 2009. Mutual Fund portfolio creation using Industry Concentration. *The ICFAI Journal of Management Research* VIII(3):7–20. <http://www.morningstar.in/featured-reports.aspx>
- Gupta, O.P., and Amitabh Gupta. 2004. Performance evaluation of Select Indian Mutual Fund Schemes: An empirical study. *The ICFAI Journal of Applied Finance* 81–97.
- Jayadeve, M. 1996. Mutual Fund performance: An analysis of monthly returns. *Finance India* X (1, March): 73–84.

- Jensen, Michael C. 1967. The Performance of Mutual Funds in the period 1945–1964. *Journal of Finance* 23 (2): 389–416.
- Jensen Michael, C. 1968. The performance of Mutual Funds in the period 1945–1964. *Journal of Finance* 23: 389–416.
- Kajshmi, N., Malabika Deo, and B. Murugesan. 2008. Performance of the Indian Mutual Funds; A study with special reference to growth schemes. *Asia Pacific Business Review* IV (3): 75–81.
- Khurshid, S.M.Z., Rohit, and G.P. Sing. 2009. Level and trends of competition among the mutual funds in India. *Research Journal of Business Management* 3 (2): 47–67.
- Lintner, John. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics* 47 (1): 13–37.
- Massa, Massimo, and LeeXhang. 2008. The effects of organizational structure on asset management. <http://ssrn.com/abstract=1328189>
- Mc Donald, John G. 1974. *Objectives and performance of Mutual Funds 1960–1967*, 311–333. June: Journal of Financial and Quantitative Analysis.
- Modigliani, Franco, and Leah Modigliani. 1997. Risk adjusted performance. *Journal of Portfolio Management*: 45–54.
- Obeid, Alexander T. 2004. A modified approach for Risk-adjusted performance attribution. *Financial Markets and Portfolio Management* 18 (3): 285–305.
- Redmand, Arnold L., N.S. Gullett, and Herman Manakyan. 2000. The Performance of Global and International Mutual Funds. *Journal of Financial and Strategic Decisions* 13 (1, Spring): 75–85.
- Sadhak, H. 2003. *Mutual Funds in India: Marketing strategies and investment practices*. New Delhi: Response Books.
- Sharpe, William F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 225–242.
- . 1966. Mutual fund performance. *Journal of Business* 39: 119–138.
- Sidana, Gajendra. 2007. Classifying mutual funds in India: Some results from clustering. *Indian Journal of Economics and Business* II (2).
- Smith, Daivd M. 2009. ‘The economics of mutual funds’, Chapter-3 of forthcoming. In *A Companion to Mutual Funds*, ed. John A. Haslem. USA: Wiley.
- Smith, and Tito. 1969. Risk-Return measures of post-portfolio performance. *Journal of Financial and Quantitative Analysis* 4: 449–471.
- Sondhi, H.J. 2007. *Financial performance of equity mutual funds in India*. New Delhi: Deep & Deep.
- Statman, M. 2000. Socially responsible mutual funds. *Financial Analysts Journal* 56: 30–38.
- Subha, M.V., and S. Bharati Jaya. 2007. An empirical study on the performance of select mutual fund schemes in India. *Journal of Contemporary Research in Management* I (1).
- Subash Chander and Japal Singh. 2004. Performance of mutual funds in India: An empirical evidence. *The ICFAI Journal of Applied Finance* pp. 45–63.
- Sudhakar, A., and K. Sasi Kumar. 2005. Performance evaluation of mutual funds: A case study. *Southern Economist*, November, pp. 19–23.
- Treynor, Jack L. 1965. How to rate management of investment funds. *Harvard Business Review* 43: 63–75.
- Treynor Jack, L., and Kay K. Mazuy. 1966. Can mutual funds outguess the markets. *Harvard Business Review* 44: 131–136.
- Treynor, Jack L., and Kay K. Mazuy. 1966. Can mutual funds outguess the markets. *Harvard Business Review* 44: 131–136.
- Volkman, D.A., and M.E. Wohar. 1995. Determinants of persistence in relative performance of mutual funds. *Journal of Financial Research* 18: 415–430.
- Wahal, Sunil, and Alber (Yan) Wang. 2010. Competition among mutual funds. *Journal of Financial Economics*, March. <http://ssrn.com/abstract=1130822>
- Yan, Y. 1999. Measuring the timing ability of mutual fund Managers. *Annals of Operations Research* 233–243.



Structural Credit Risk Models: Endogenous Versus Exogenous Default

56

Michael B. Imerman

Abstract

This chapter reviews structural credit risk models. Special emphasis is on the distinction between endogenous default versus exogenous default and the economic implications of the different assumptions. It is argued that models with endogenous default provide more insight into the default process. On the other hand, assuming exogenous default gives the flexibility to include certain features that are observed in actual credit markets.

Keywords

Bond pricing · Credit spreads · Default probability · Structural credit risk models

56.1 Introduction

This chapter surveys the literature on structural credit risk models with emphasis on the distinction between models with endogenous default versus those with exogenous default. It is argued that models with exogenous default may be internally inconsistent and even biased. With exogenous default an arbitrary default barrier is assumed and, therefore, default probabilities can be manipulated by modifying the assumptions related the barrier. These types of problems, although not entirely

M. B. Imerman (✉)
Princeton University, Princeton, NJ, USA

Table 1 Select structural credit risk models and their features

Model	Default	Recovery	Interest rates
Black and Cox (1976)	Exogenous	Fixed fraction of present value of face value of debt	Deterministic
	Exponential barrier		
Longstaff and Schwartz (1995)	Exogenous	Allows for violations of ARP	Stochastic
	Flat barrier		
Collin-Dufresne and Goldstein (2001)	Exogenous	Same as Longstaff- Schwartz	Stochastic
	Mean-reverting barrier (stationary leverage)		
Geske (1977)	Endogenous	Asset value at default	Deterministic
Geske and Johnson (1984)	Function of liabilities		
Leland (1994)	Endogenous	Fixed proportion of asset value	Deterministic
Leland and Toft (1996)	Flat barrier		

eliminated, are less likely to occur with endogenous models.¹ Endogenous models, on the other hand, solve for the point at which the firm will default under optimal conditions. They tend to provide more insight in terms of the default process (when firms default) and allow for meaningful comparative statics (how default probabilities change with respect to the underlying variables). However, that is not to say structural models with exogenous default are useless; in fact, exogenous default models still make many contributions to the analysis of credit risk, just in different capacities. It will be shown that imposing exogenous default allows for interesting real-world features to be incorporated into the model, including deviations from the absolute priority rule and mean-reverting leverage ratios.

Structural credit risk models were pioneered by the seminal works of Black and Scholes (1973) and Merton (1974), who noted that the equity is essentially a call option on the firm's assets when the firm has debt in its capital structure. Though the insight was groundbreaking, the model is overly simple to the point where using it to model real-world credit risk becomes rather implausible. Jones et al. (1984) confirmed this empirically where it was shown that the basic Black-Scholes-Merton (BSM) model generates credit spreads that are smaller than are actually observed.² As such, in the decades that followed researchers have developed, implemented, and tested extensions to the basic BSM model utilizing the contingent-claims framework but incorporating more realistic features of corporate debt. In fact, the literature on structural credit risk models has become so extensive that a comprehensive survey could be a book in and of itself. Instead, this chapter focuses on a handful of popular structural credit risk models: two with endogenous default and three with exogenous

¹Of course, the parameters in a structural credit risk model can still be "tweaked" to obtain desirable results, but the risk of modeler manipulation is present with any financial model.

²The result is that the BSM model *overprices* risky debt.

default (in addition to the BSM and KMV). These models and their main features are outlined in Table 1.

The remainder of the chapter is organized as follows. In Sect. 56.2 two basic structural models are discussed: the original Black-Scholes-Merton model and the Moody's-KMV model as used in the literature.³ Section 56.3 discusses three structural credit risk models with exogenous default and Sect. 56.3.2 discusses two structural credit risk models with endogenous default. Finally, Sect. 56.4 concludes.

56.2 The Basic Model

56.2.1 Black-Scholes-Merton (BSM)

One of the first papers to discuss pricing debt and other corporate liabilities as contingent claims on the underlying firm value was Black and Scholes (1973). Although this paper is usually cited for its contribution to option pricing theory, the last few sections deal with the more subtle problem of accurate valuation of corporate securities, including common stock and corporate bonds.

The concept of option pricing lends itself quite nicely to the valuation of corporate securities. When a firm's capital structure contains debt, the equity can be seen as a long call option on the firm's asset value. When the debt is in the form of a discount bond paying F at maturity time T , then the equity is worth $\text{Max}\{V_T - F, 0\}$ at maturity, where V_T is the terminal value of the firm's assets at maturity time T . This is exactly the same payoff at expiration of a long European call option with strike price F . The equilibrium price of a contingent claim paying $V_T - F$ if $V_T > F$, and nothing otherwise, is found by solving a partial differential equation that describes the functional relationship between the contingent claim and the underlying asset subject to the boundary conditions, $E(V_T, T) = V_T - F$, if $V_T > F$; 0, if $V_T \leq F$. When the asset value evolves according to a Geometric Brownian Motion, then the intertemporal equilibrium value of equity is the value of the contingent claim E_t as defined by the well-known Black-Scholes formula:

$$E_t = V_t N(d_1) - F e^{-r(T-t)} N(d_2) \quad (1)$$

where, $d_1 = \frac{\ln(\frac{V_t}{F}) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$; $d_2 = \frac{\ln(\frac{V_t}{F}) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$.

The function $N(\cdot)$ is the cumulative normal distribution function.

Since the sum of the market value of debt and the market value of equity must equal the market value of the assets, then by definition the intertemporal equilibrium

³The actual model that Moody's-KMV uses in their analysis is proprietary and remains a "black box" to the public. However, a variation of the Moody's-KMV model is used widely in the academic and practitioner literature; it is that variation will be discussed in this chapter.

value of the debt is $V_t - E_t$. This implies that the bondholders essentially own the firm's assets and sell a call option on the assets to the stockholders. Since, in this basic setup, the debt is a pure discount bond, the bondholders pay less than the face value F at time 0, where the discount is $F - (V_0 - E_0)$. At maturity, time T , the bondholders either get the face value of the debt (F) or, if the value of the assets is not high enough to cover the face value, then the firm is "bankrupt" and the bondholders own the assets at value V_T (i.e. the call option is not exercised). The difference between the amount paid at time 0 and F consists of both the time value and the premium due to default risk.

There is another way to look at the debt of a firm, arising from the put-call parity identity first explicated by Stoll (1969). Put-call parity links four different markets: the call option market, the put option market, the market for the underlying asset, and the [risk-free] money market. It states that the value of a portfolio consisting of a long call option (with strike price X and maturity T) and cash equal to the present value of the strike price must be equal to a portfolio consisting of a long position in the underlying asset and a long put option (also with strike price X and maturity T). In symbols, this says $c_t + PV(X) = S_t + p_t$. When E_t is a long call option on the firm value (V_T) with strike price equal to the face value of debt (F), we get $E_t + PV(F) = V_t + p_t$. Rearranging, we see that the value of the risky discount bond described above is $V_t - E_t = PV(F) - p_t$. This decomposes the risky bond into two parts: the present value of the principal that is lent ($PV(F)$) and a short put option on the firm's assets ($-p_t$). The former is the pure time value at the risk-free rate; the latter is the default risk premium. When a firm's capital structure consists of just equity and one class of non-coupon paying debt, the rearranged put-call parity identity can be thought of as a capital markets equilibrium condition that links the value of the firm's assets, the value of the common stock, and the value of risky debt (which consists of pure time value and value attributed to the risk of default).

Merton (1974) adds to the Black and Scholes (1973) analysis with an entire paper dedicated to the idea of applying option pricing methodology to value risky debt. The paper seeks to quantify the increase or decrease in the prices of bonds that result from changes in the probability that an issuer will not be able to satisfy their debt obligations (i.e. default risk). As the paper focuses on the "risk structure of interest rates," Merton attempts to isolate the impact of default risk by assuming a flat term structure with a constant risk free rate. However, this assumption is not only unrealistic but is also quite problematic and may explain why the basic BSM structural model tends to overprice debt in reality.

The Merton (1974) paper is much more mathematically rigorous than the Black and Scholes (1973) paper and contains complete comparative statics of the debt value as a function of the underlying firm value, the degree of leverage ("quasi" debt-to-firm value ratio), the volatility of the firm, and the time-to-maturity. The same comparative statics are reproduced using the default risk premium (defined as the difference between the yield-to-maturity on the risky bond and the risk-free rate).

Aside from the flat yield curve assumption, there are several other simplifications in the Merton (1974) model that make it impractical and unrealistic. These simplifications and the problems resulting from them provide much of the impetus for the extensions that would emerge over the following years. The basic Merton (1974) model assumes that the firm only has one issue of debt outstanding and that the issue consists of non-coupon paying debt. This is probably unrealistic as many firms have several different debt issuances in their capital structure at a given time and quite often they are of different priority and maturity. Most bonds also pay coupons, either fixed or floating rate.⁴ Next section, a practical implementation of the BSM structural credit risk model – the Moody’s-KMV model – is discussed. The Moody’s-KMV model is also a single-period model, but it incorporates more information about the liability structure (short-term and long-term debt) and is often used to construct an ad-hoc term structure of default probabilities.

56.2.2 Moody’s-KMV (MKMV)

The Moody’s-KMV (MKMV) structural credit risk model is a practical implementation of the original BSM model with certain modifications to reflect empirically observed characteristics of default. The model was developed by KMV corporation which was subsequently acquired by Moody’s Corporate in 2002. The true KMV model used for commercial purposes is proprietary and can be considered a “black box”; however, a well-known version described in the technical document by Crosbie and Bohn (2003) is commonly used in industry and in academic studies.⁵

The MKMV model has exogenous default, but unlike the BSM model, the default barrier is not assumed to be the face value of all debt. Rather Crosbie and Bohn (2003) note that, empirically, default typically occurs at values *below* the total face value of debt. Since short-term debt represents a more immediate constraint, that is usually taken at 100% of face value for the barrier; to this, a fraction of the outstanding long-term debt is then added. Letting K_{MKMV} denote the default barrier for the Moody’s-KMV model, it is generally accepted in the literature to specify it as:

$$K_{KMV} = ST + \frac{1}{2}LT \quad (2)$$

⁴It should be noted that at the end of the paper Merton (1974) does in fact address the pricing of coupon paying bonds; however, the valuation is significantly more complicated and very impractical to apply in practice.

⁵In fact, quite often this version of the KMV structural model is used as a benchmark default model against which others are compared in the academic literature. See, for instance, Bharath and Shumway (2008), Leland (2004), etc.

where ST denotes the total face value of short-term liabilities and LT denotes the total face value of long-term liabilities. This default barrier is then used as the strike price in the Black-Scholes formula for computing the European call option value. The other inputs are: the market value of the firm's assets (V), the asset volatility (σ_V), the risk-free rate (r), and the default horizon (T). Now, however, the modeler is faced with an "inverse problem"—specifically, two of the inputs are unobservable (asset value and asset volatility) whereas the output (equity value) is observable. There is a clever solution to this problem. From Ito's Lemma we can specify the relationship between equity volatility (observable) and asset volatility (unobservable), in closed-form. In general, this relationship is⁶:

$$\sigma_E E = \sigma_V V \frac{\partial E}{\partial V} \quad (3)$$

In the Black-Scholes model, it is well known that the partial derivative of the call option value with respect to the underlying asset value is known to be $N(d_1)$ from Eq. 1, the option "Delta," this can be plugged into Eq. 3 which results in two equations in two unknowns:

$$\begin{aligned} E &= BS_{Call}(V, \sigma_V, K_{MKMV}, T, r) \\ \sigma_E &= \frac{V}{E} N(d_1) \sigma_V \end{aligned} \quad (4)$$

where $BS_{Call}(V, \sigma_V, K_{MKMV}, T, r)$ is the function given in Eq. 1. In order to solve the system of equations given in Eq. 4 an iterative numerical procedure has to be used. Typically one selects initial values for V and σ_V , say the book value of assets and some fraction of the empirical equity volatility. Then the algorithm will go back and forth between the two equations seeking to minimize the distance between the model outputs for E and σ_E with the observed values for the market value of equity and empirical equity volatility, respectively.

Once an appropriate default horizon has been chosen, say $T = 1$ year, the MKMV model can then be used to compute the probability that the particular firm will default given the current liability structure (i.e. book value of short-term and long-term debts), the market value of the firm's equity, an estimate of the equity volatility, and the risk-free interest rate. This is the probability that a European call option on the firm's assets expires out-of-the-money where the strike price is equal to the MKMV default barrier (from Eq. 2) and the option expiration is equal to the default horizon. The default probability is:

$$\Pr\{V_T \leq K_{MKMV}\} = 1 - N(d_2) = N(-d_2) \quad (5)$$

where $N(d_2)$ is defined in the valuation expression in Eq. 1.

⁶This is true not only for Black-Scholes options but for any contingent claims.

Another metric from the MKMV model that has become popular is the “Distance-to-Default” or DD. The Distance-to-Default shows, for a given horizon T , how many standard deviations away from the default barrier the expected asset value will be T years out. Since it is assumed that asset value follows a Geometric Brownian Motion, the Distance-to-Default is:

$$DD_T = \frac{\ln\left(\frac{V_0}{K_{MKMV}}\right) + \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}. \quad (6)$$

This happens to be d_2 from Eq. 1. Consequently, we can rewrite the default probability in Eq. 5 as $N(-DD_T)$.

Both the original BSM structural credit risk model and the MKMV structural credit risk model have exogenous default, since the default point is defined exclusively in terms of the face values of debts. In the next section, three extensions to this basic structural framework are discussed. All three also have exogenous default, but default can occur at any time in the form of a Markov hitting time.

56.3 Exogenous Default Models

56.3.1 Black and Cox (BC)

Black and Cox (1976) extend the basic structural model to incorporate some specific restrictions that may appear in the bond indenture. These restrictions can affect the conditions under which a firm defaults and, therefore, will have an impact on the pricing of risky debt. They begin with assumptions similar to those made in Merton (1974) leading to the general partial differential valuation equation. The authors point out that the partial differential equation already implies certain restrictions on investment policy, financing policy, and payouts. In this paper, Black and Cox (1976) examine “safety covenants”—contractual provisions that stipulate the firm-specific conditions that force restructuring. There may or may not be an upper boundary on firm value, above which the firm is to call debt. More importantly there is a lower boundary on firm value, below which the firm is “bankrupt”; here the debt holders take control of the firm’s assets. Black and Cox (1976) were the first to consider this notion of a default barrier; they note that it may be either exogenous, a given value specified outside the model, or endogenous, where it is actually solved for in the model. Endogenous barriers represent an optimal decision problem, whereas exogenous barriers are specified in the contract. Since the Black and Cox (1976) model deals with safety covenants, the barriers are exogenous. They note that the barrier may take on different forms depending on the specific type of covenant that is being modeled. Perhaps the most well-recognized form is the exponential barrier where bankruptcy is triggered when the asset value V_t falls to the level $K(t) = Ce^{-\gamma(T-t)}$, where C is a constant that may be a function of

the initial face value of the debt issued, T is the maturity, and γ is the barrier growth rate.

Assuming that the asset value evolves according to a Geometric Brownian Motion, then this is a Gaussian first passage problem. The default probability, or the probability that the asset value hits the barrier by time τ is

$$1 - N\left(\frac{\ln\left(\frac{V}{Ce^{-\gamma(T-\tau)}}\right) + \left(r - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}}\right) + \left(\frac{V}{Ce^{-\gamma(T-\tau)}}\right)^{1-2\left(\frac{r-\gamma}{\sigma^2}\right)} N\left(\frac{\ln\left(\frac{Ce^{-\gamma(T-t)}}{V}\right) + \left(r - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}}\right) \quad (7)$$

where $N(\cdot)$ represents the cumulative standard normal distribution function.

While arbitrary values can be chosen for C and γ – thereby rendering the default barrier arbitrary – Black and Cox (1976) note that a reasonable and “interesting” choice is to set C equal to ρF and γ equal to the risk-free rate, where $0 \leq \rho \leq 1$. This implies that the firm will default when asset value equals a specified fraction of the present value of the principal amount borrowed (i.e. the face value). As a result, this imposes an exogenous recovery assumption with a fixed recovery rate of ρ . In this case, $K(t) = \rho Fe^{-r(T-t)}$.

The market value of the debt, D_t , can be found as the discounted expected value of the payoffs, which are given by the boundary conditions at maturity and default, respectively:

$$D_T = \min\{V_T, F\} \quad (8)$$

$$D_\tau = \rho Fe^{-r(T-\tau)} \quad (9)$$

The value at default, given by Eq. 9, assumes the recovery rate discussed above. This leads to the following valuation formula for debt:

$$D(t) = Fe^{-r(T-t)}[N(z_1) - y^{2\theta-2}N(z_2)] + Ve^{-a(T-t)}[N(z_3) + y^{2\theta}N(z_4) + y^{\theta+\zeta}e^{a(T-t)}N(z_5) + y^{\theta-\zeta}e^{a(T-t)}N(z_6) - y^{\theta-\eta}N(z_7) - y^{\theta-\eta}N(z_8)] \quad (10)$$

where

$$y = \frac{Ce^{-\gamma(T-t)}}{V_t}$$

$$\theta = \frac{\left(r - a - \gamma + \frac{\sigma^2}{2}\right)}{\sigma^2}$$

$$\delta = \left(r - a - \gamma - \frac{\sigma^2}{2}\right) + 2\sigma^2(r - \gamma)$$

$$\zeta = \frac{\sqrt{\delta}}{\sigma^2}$$

$$\eta = \frac{\sqrt{\delta - 2\sigma^2 a}}{\sigma^2}$$

$$z_1 = \frac{\ln\left(\frac{V_t}{F}\right) + \left(r - a - \frac{\sigma^2}{2}\right)(T - t)}{\sqrt{\sigma^2(T - t)}}$$

$$z_2 = \frac{\ln\left(\frac{V_t}{F}\right) + 2\ln y\left(r - a - \frac{\sigma^2}{2}\right)(T - t)}{\sqrt{\sigma^2(T - t)}}$$

$$z_3 = \frac{\ln\left(\frac{F}{V_t}\right) - \left(r - a + \frac{\sigma^2}{2}\right)(T - t)}{\sqrt{\sigma^2(T - t)}}$$

$$z_4 = \frac{\ln\left(\frac{V_t}{F}\right) + 2\ln y\left(r - a + \frac{\sigma^2}{2}\right)(T - t)}{\sqrt{\sigma^2(T - t)}}$$

$$z_5 = \frac{\ln y + \zeta\sigma^2(T - t)}{\sqrt{\sigma^2(T - t)}}$$

$$z_6 = \frac{\ln y - \zeta\sigma^2(T - t)}{\sqrt{\sigma^2(T - t)}}$$

$$z_7 = \frac{\ln y + \eta\sigma^2(T - t)}{\sqrt{\sigma^2(T - t)}}$$

$$z_8 = \frac{\ln y - \eta\sigma^2(T - t)}{\sqrt{\sigma^2(T - t)}}$$

and a , which was not previously defined, is the continuous dividend rate; if no dividends are paid – as was assumed earlier – then $a = 0$ and Eq. 10 simplifies significantly.

The Black and Cox (1976) model offers several advantages. First, it allows default to occur at anytime between now and maturity. However, the fact that default usually occurs at discrete points in time (i.e. when payments are due to creditors) can be viewed as a drawback, which is addressed by other models (see, e.g., the Geske compound option model in Sect. 56.3.3). Another advantage is that the model has a closed-form solution and is easy to implement and, in fact, serves as the template for many of the models that are used in industry. The largest disadvantage is that it does not work when bankruptcy is endogenous.

56.3.2 Longstaff and Schwartz (LS)

Longstaff and Schwartz (1995) essentially extend the Black and Cox (1976) model. It incorporates stochastic interest rates by including the Vasicek (1977) mean-reverting term structure dynamics so that the model is a two-factor Markov process. There is a constant, or flat, exogenous default barrier set equal to K . The fact that K is fixed for the life of the firm implies a stationary capital structure. This is achieved by assuming that coupons and principal payments are funded by issuing new debt. There is also an exogenous recovery assumption where, upon default, creditors receive a fixed percentage of face value. This actually allows for violations of the Absolute Priority Rule (APR) which is a nice feature of the model. Longstaff and Schwartz (1995) derive elegant, closed-form solutions for pricing both fixed and floating rate debt with various maturities using the structural credit risk model.

The LS model begins, as most structural credit risk models do, with the assumption that the asset value, V_t , follows a Geometric Brownian Motion. Given the flat barrier, default is modeled as a first passage event. As long as $V_t > K$, then the firm remains solvent; however, once $V_t \leq K$ financial distress is triggered. When the firm is distressed there is assumed to be some reorganization that occurs, which results in debt holders receiving a fraction, $w \leq 1$, of the face value. Since w is the “write-down” on the debt, the recovery rate would be $R = 1 - w$. The authors note that, when a firm has various classes of debt outstanding, there may be different values of w representing differential recovery rates (for instance w_{sr} , and w_{sub} , for Senior and Subordinated debt, respectively).

The value of a fixed-rate risky discount bond, subject to both interest rate and default risk, is a function of the following variables: $X = \frac{V}{K}$, r , and T , which represent a measure of distance from the default barrier, the [stochastic] interest rate, and the time-to-maturity, respectively; as well as the parameters: α , β , η , ρ , σ , and w , where the first three are interest rate parameters from the Vasicek (1977) model, ρ is the correlation between the Brownian Motion that drives the interest rate dynamics and the Brownian Motion that drives the asset value dynamics, σ is the asset volatility, and w is the “write-down” as described in the last paragraph.

Define the face value of the bond with maturity T to be $F(r, T)$ and let $D(X, r, T)$ be the price of a risky discount bond:

$$D(X, r, T) = F(r, T) - wF(r, T)Q(X, r, T) \quad (11)$$

where $Q(X, r, T)$ is the risk-neutral default probability defined as the limit where $n \rightarrow \infty$ of

$$\begin{aligned} Q(X, r, T, n) &= \sum_{i=1}^n q_i \\ q_1 &= N(a_1) \\ q_i &= N(a_i) - \sum_{j=1}^{i-1} q_j N(b_{ij}), i = 2, 3, \dots, n \\ a_i &= \frac{-\ln X - M\left(\frac{iT}{n, T}\right)}{\sqrt{S\left(\frac{iT}{n}\right)}} \\ b_{ij} &= \frac{M\left(\frac{jT}{n, T}\right) - M\left(\frac{iT}{n, T}\right)}{\sqrt{S\left(\frac{iT}{n}\right) - S\left(\frac{jT}{n}\right)}} \end{aligned}$$

where $N(\cdot)$ is the cumulative standard normal distribution function and

$$\begin{aligned} M(t, T) &= \left(\frac{\alpha - \rho\sigma\eta}{\beta} - \frac{\eta^2}{\beta^2} - \frac{\sigma^2}{2}\right)t + \left(\frac{\rho\sigma\eta}{\beta^2} + \frac{\eta^2}{2\beta^3}\right)e^{-\beta T}(e^{\beta t} - 1) \\ &\quad + \left(\frac{r}{\beta} - \frac{\alpha}{\beta^2} + \frac{\eta^2}{\beta^2}\right)(1 - e^{-\beta t}) - \left(\frac{\eta^2}{2\beta^3}\right)e^{-\beta T}(1 - e^{-\beta t}), \\ S(t) &= \left(\frac{\rho\sigma\eta}{\beta} + \frac{\eta^2}{\beta^2} + \sigma^2\right)t - \left(\frac{\rho\sigma\eta}{\beta^2} + \frac{2\eta^2}{\beta^3}\right)(1 - e^{-\beta t}) \\ &\quad + \left(\frac{\eta^2}{2\beta^3}\right)(1 - e^{-2\beta t}) \end{aligned}$$

The LS model can be used to price coupon-paying bonds as a portfolio of these risky discount bonds.

Floating-rate payments are valued as explicit functions of the stochastic interest rate

$$C(X, r, \tau, T) = D(X, r, T)R(r, \tau, T) - wF(r, T)G(D, r, \tau, T) \quad (12)$$

where

$$\begin{aligned} R(r, \tau, T) &= re^{-\beta\tau} + \left(\frac{\alpha}{\beta} - \frac{\eta^2}{\beta^2}\right)(1 - e^{-\beta\tau}) + \left(\frac{\eta^2}{2\beta^2}\right)e^{-\beta T} \\ (e^{\beta\tau} - e^{-\beta\tau})G(X, r, \tau, T, n) &= \sum_{i=1}^n q_i \frac{K\left(\tau, \frac{iT}{n}\right)}{S\left(\frac{iT}{n}\right)} M\left(\frac{iT}{n}, T\right) \\ K(\tau, t) &= \left(\frac{\rho\sigma\eta}{\beta} + \frac{\eta^2}{\beta^2}\right)e^{-\beta\tau}(e^{\beta \min\{\tau, t\}} - 1) - \left(\frac{\eta^2}{2\beta^2}\right)e^{-\beta\tau}e^{-\beta\tau}(e^{2\beta \min\{\tau, t\}} - 1) \end{aligned}$$

and $G(X, r, \tau, T)$ is the limit of $G(X, r, \tau, T, n)$ as $n \rightarrow \infty$. Floating-rate coupon bonds can be valued as the sum of floating rate cash flows in Eq. 12 plus the value of the principal from Eq. 11.

Longstaff and Schwartz (1995) are able to perform some interesting empirical analysis using their model. Many of the insights stem from correlations between credit spreads, interest rates, and equity returns. For instance, when regressing changes in credit spreads (from Moody's bond yield data) on changes on Treasury yields and returns on an appropriate equity index, they find that the coefficient on both interest rates and equity returns are negative, although significance varies across industries and credit ratings. This is consistent with the LS model, which implies that credit spreads tighten as interest rates increase and credit spreads are negatively related to equity returns. The latter point is very intuitive from a credit risk perspective: holding all else constant, an increase in equity corresponds to an increase in asset value which results in lower leverage and lower default probabilities. Lower default probabilities should be reflected in a smaller default risk premium and tighter credit spreads. The key innovation, and indeed one of the main contributions, lies in the impact of stochastic interest rates on credit spreads and bond prices.

56.3.3 Collin-Dufresne and Goldstein (CDG)

Collin-Dufresne and Goldstein (2001) claim that target leverage ratios affect credit spreads. They develop a structural model of debt that allows for mean-reverting leverage and show that this model produces credit spreads that are more in line with those observed empirically. The Collin-Dufresne and Goldstein (2001)—referred to as CDG—model is a direct extension of the Longstaff and Schwartz (1995) model that includes a stochastic default boundary and an option to add more debt. This option adds value and seems to be reflected in credit spreads.

Most structural models assume a static capital structure of the firm; that is to say that the firm cannot and does not act to dynamically add or reduce debt.^{7, 8} However, this behavior is not representative of how firms act. Collin-Dufresne and Goldstein (2001) argue that the valuation of debt must take into account not only the firm's current leverage but also the option to increase this leverage in the future; ignoring this option can result in the mispricing of debt within the structural model. Specifically, the authors claim that the firm's option to issue more debt should increase credit spreads on current outstanding debt. The intuition for why is rather straightforward: a firm not constrained with respect to future capital structure is likely to take on more debt in the future if conditions are favorable. This possibility makes default more likely in the future than if capital structure was restricted and, therefore, debt holders are going to price the current outstanding debt to account for this added risk. There is empirical evidence that leverage ratios are stationary, which implies that as firm value rises (falls), the firm will issue (repurchase) debt. Thus, a structural credit risk model should be able to account for these changes in the default barrier, moving up and down with leverage. However, we do not know when these changes will occur and so leverage, and consequently the barrier, is modeled as a stochastic process itself. The CDG model is therefore an exogenous barrier model, like the other models discussed thus far. The difference is that the barrier is not flat or following some deterministic function, but itself is driven by randomness. The mechanisms for this process will be discussed next.

Assuming that the asset value dynamics follow a Geometric Brownian Motion, then asset values, V_t , are lognormally distributed. In the Collin-Dufresne and Goldstein (2001) paper everything is converted to log-values for convenience. Therefore, the main state variable is going to be $y_t = \log V_t$. Next define the "log-default threshold," k_t , which represents a log-transformation of debt at time t . Then, the log-leverage is $l_t = k_t - y_t$. The log-leverage follows a mean-reverting Markov process, where the stochastic component comes from the Wiener process driving y_t ; the mean-reversion comes from the fact that k_t adjusts based on its proximity to y_t (when k_t is high the barrier decreases and when k_t is low the barrier increases). Note that there are two Brownian Motions in the model: one for the asset value and another for the interest rate. Since the mean-reverting leverage process is actually driven by the same random factor as the asset value process, the CDG model can be thought of as a "two-and-a-half factor" structural credit risk model. The target leverage ratio is actually a decreasing function of the current interest rate, which reflects that firms increase leverage when interest rates are low and decrease leverage when interest rates are high. The model yields several other results that appear to be consistent with empirical evidence about credit markets. Whereas previous structural

⁷The Geske (1977) compound option structural credit risk model new equity is raised to pay down outstanding debt. This is not a dynamic rebalancing of the capital structure, but rather a systematic reduction in the amount of leverage in a manner that is typically not seen in practice. This will be discussed in detail in Sect. 56.3.3.

⁸Goldstein et al. (2001) do in fact develop a structural model for dynamic capital structure.

credit risk models have performed poorly in generating credit spreads and for bond pricing (see, e.g., Jones et al. (1984) and Eomet et al. (2004)), the CDG model provides a theoretical basis for explaining these discrepancies. In fact, recent work by Flannery et al. (2011) empirically confirm that changes in credit spreads reflect not only the current leverage ratio but also the market's expectations about *future* leverage ratios. This represents a major breakthrough towards understanding the impact of capital structure decisions on the dynamics of bond prices, credit spreads, and default probabilities. However, as will be discussed in the next section, the incorporation of endogenous default into a structural credit risk model provides a more integrative way to unify capital structure decisions and the various dimensions of credit risk.

56.4 Endogenous Default Models

56.4.1 Geske (G)

Geske (1977) and Geske and Johnson (1984) value corporate securities (both debt and equity) as *compound options*. The compound option approach has endogenous default and, while distinct from the typical barrier/first-passage structural models, can be thought of as a discrete barrier. This approach was originally proposed by Geske (1977), who noted that when there are multiple cash flows, in the form of coupon payments and/or several classes of debt outstanding, then equity is a *compound* call option on the firm's assets.⁹ A compound call option is a call option on a call option. Each cash flow is therefore a strike price for the compound option. Whenever a payment is due to bondholders, the firm's shareholders must decide whether or not to exercise the compound call option. Making the payment represents exercising the compound call option, thereby agreeing to continue holding a sequence of contingent claims on the firm's assets. When there are two cash flows the model solution is known in closed-form using the compound option formulae from Geske (1979). This set-up is useful for valuing risky debt when the liability structure is bifurcated into short-term and long-term debt or junior and senior debt. Geske and Johnson (1984) appropriately specify the default conditions when there are two distinct classes of debt, which leads to the correct closed-form expression for the intertemporal values of equity, junior, and senior debt.

Let the value of the firm's assets follows a Geometric Brownian Motion satisfying the stochastic differential equation

$$dV_t = r_t V_t dt + \sigma_V V_t dW_t \quad (13)$$

⁹The idea of modeling coupon-paying debt as a compound option was actually proposed by Black and Scholes (1973) [see pp. 651–652]; however, compound options cannot be priced by the standard Black-Scholes valuation formula.

where dW_t represents a standard Brownian increment under the risk-neutral probability measure. Then the valuation of the corporate securities in this framework are given by the expectation of the discounted future payoffs under the risk-neutral measure. For equity that is

$$\mathbb{E}[e^{-rT_1} \max \{E_{T_1} - F_1, 0\}] \quad (14)$$

where F_1 is the face value of the senior debt with maturity time T_1 . E_{T_1} is actually the Black-Scholes value of the European call option expiring at time $T_2 > T_1$ (T_2 is the maturity time for the junior debt which has face value F_2). The value E_{T_1} is not known at the current time $t = 0$, and plays a key role in determining the endogenous default point in the compound option structural model. The expectation given in Eq. 14 is equal to

$$E_{t=0} = V_{t=0} N_2(h_1^+, h_2^+; \rho) - F_2 e^{rT_2} N_2(h_1^-, h_2^-; \rho) - F_1 e^{-rT_1} N(h_1^-) \quad (15)$$

where

$$\begin{aligned} h_1^+ &= \frac{\ln\left(\frac{V_t}{V}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau_1}{\sigma\sqrt{\tau_1}}, \\ h_1^- &= \frac{\ln\left(\frac{V_t}{V}\right) + \left(r - \frac{\sigma^2}{2}\right)\tau_1}{\sigma\sqrt{\tau_1}} = h_1^+ - \sigma\sqrt{\tau_1}, \\ h_2^+ &= \frac{\ln\left(\frac{V_t}{F_2}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau_2}{\sigma\sqrt{\tau_2}}, \\ h_2^- &= \frac{\ln\left(\frac{V_t}{F_2}\right) + \left(r - \frac{\sigma^2}{2}\right)\tau_2}{\sigma\sqrt{\tau_2}} = h_2^+ - \sigma\sqrt{\tau_2}, \end{aligned}$$

$N(\cdot)$ denotes the cumulative standard normal distribution,

$N_2(\cdot)$ denotes the cumulative bivariate standard normal distribution, and correlation $\rho = \sqrt{\frac{\tau_1}{\tau_2}}$ which follows from the properties of Brownian Motion.

$\bar{V}$ is the endogenous default boundary.

Now, the senior debt has face value F_1 and is to be paid at time T_1 . Holders of the senior debt will receive either the total amount they are promised (i.e. no default) or, in the event there is a default, the assets will be liquidate and they will receive the market value of the assets at time T_1 . Thus, the value of the senior debt at time $t = 0$ is given by the expectation

$$\mathbb{E}[e^{-rT_1} \min \{V_{T_1}, F_1\}]. \quad (16)$$

It is fairly straightforward to show that the expectation given in Eq. 16 is equal to

$$S_0 = V_0[1 - N(k^+)] + F_1 e^{-rT_1} N(k^-) \quad (17)$$

$$\text{where } k^+ = \frac{\ln\left(\frac{V_t}{F_1}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau_1}{\sigma\sqrt{\tau_1}} \text{ and } k^- = \frac{\ln\left(\frac{V_t}{F_1}\right) + \left(r - \frac{\sigma^2}{2}\right)\tau_1}{\sigma\sqrt{\tau_1}} = k^+ - \sigma\sqrt{\tau_1}.$$

Then, finally, the value of the junior debt at time $t = 0$ is given by the expectation:

$$\mathbb{E}\left[e^{-rT_1} \left(\left(J_{T_1} \cdot \mathbb{I}_{\{V_{T_1} > \bar{V}\}} \right) + \left(\max\{V_{T_1} - F_1, 0\} \cdot \mathbb{I}_{\{V_{T_1} \leq \bar{V}\}} \right) \right) \right] \quad (18)$$

where $\mathbb{I}_{\{V > \bar{V}\}}$ is the indicator function that equals 1 if $\mathbb{I}_{\{V > \bar{V}\}}$ and 0 otherwise.

To obtain the closed-form solution for the junior debt, the expectation given in Eq. 18 can be evaluated directly or, since the expectation for the senior debt is considerably more straightforward, the fact that $V_t = E_t + J_t + S_t$ can be used to solve for $J_t = V_t - E_t - S_t$, at any time t . Subtracting Eqs. 15 and 17 from V_0 gives:

$$\begin{aligned} J_0 &= V_0 [N(k^+) - N_2(h_1^+, h_2^+; \rho)] + F_1 e^{-rT_1} [N(h_1^-) - N(k^-)] \\ &\quad + F_2 e^{-rT_2} N_2(h_1^-, h_2^-; \rho) \end{aligned} \quad (19)$$

where everything defined in Eqs. 15 and 17 remain the same.

When there are more than two cash flows, the compound option model is known in “quasi-closed-form” in that it can be written out, but not solved analytically. Therefore, numerical methods must be used in these instances to both value the corporate securities as well as compute default probabilities. Default is endogenously determined at every cash flow, and as a result the model is able to generate a complete term structure of default probabilities. This concept of a complete term structure of default probabilities will be elaborated on shortly.

In the compound option framework, the endogenous default condition is as follows. Every time a payment is due, a decision has to be made as to whether or not the firm should default. If the asset value is sufficiently high, then there is a non-zero equity value; under the assumption of perfect markets, new equity can be raised at the current price and the proceeds used to make the imminent payment to creditors. When asset value drops, so does the value of the equity, which is a contingent claim on the assets. If the asset value gets too low, then adequate equity cannot be raised to satisfy the next obligation and the shareholders choose to *not exercise* the compound call option (i.e. default) and creditors take control of the assets in accordance to absolute priority. Thus, default is endogenously determined as the asset value that sets the value of equity just equal to the next cash flow. If the asset value falls below this level it implies negative equity value. This condition (inability to raise capital to service debt and/or negative equity value) is similar to other models with endogenous default.

Stated more formally, at every cash flow time there exists a critical asset value, $V_{T_i} = \bar{V}$, that solves the integral equation:

$$E_{T_i}(\bar{V}) - F_i = 0. \quad (20)$$

In fact, the solution to this integral equation implies that the critical asset value is exactly equal to the face value of the next cash flow plus the present value of the expected remaining cash flows thereafter. Therefore, the discrete default barrier – constructed as the sequence of critical asset values at each cash flow – reflects the firm's ability to raise capital to satisfy the next obligation as well as the asset value capacity to cover the expectation of remaining payments.

The compound option structural model assumes that new equity is issued to pay down maturing debts. This implies that, in the Geske framework, the firm in question is expected to systematically reduce its leverage over time. This assumption may be unrealistic in modeling many firms and, indeed, a major criticism of the compound option structural model is that it imposes this financing restriction. As we will see, other structural models treat the financing of maturing debt differently, which directly impacts the shape and form of the default barrier over time.

In the compound option structural credit risk model there is a default probability associated with each and every cash flow over time. As a result the compound option model explicitly gives a complete term-structure of default probabilities. The term structure of default probabilities is not just a curve of the cumulative default probabilities, which can be computed with any barrier structural credit risk model; rather it includes both conditional and unconditional forward default probabilities (i.e. marginal and joint probabilities). In the two cash flow case all of the default probabilities are known in closed-form and will be presented below.

The first probability to define is the short-term survival probability, or the probability that at time T_1 the asset value is greater than the endogenous default point. Under the risk-neutral measure this is:

$$\Pr\{V_{T_1} > \bar{V}\} = N(h_1^-) \quad (21)$$

where $N(h_1^-)$ is as defined in the valuation equations above (see, e.g., Eq. 15).

The short-term default probability, $DP(T_1)$, can now be easily defined as $1 - N(h_1^-)$ or $N(-h_1^-)$. Since the two events – surviving past the first cash flow and defaulting at time T_1 – are mutually exclusive, the probability that the asset value at time T_1 is less than or equal to the endogenous default boundary is equal to one minus the probability given in Eq. 21.

$$\Pr\{V_{T_1} \leq \bar{V}\} = 1 - \Pr\{V_{T_1} > \bar{V}\} = 1 - N(h_1^-) = N(-h_1^-). \quad (22)$$

The next probability to define is the total survival probability, which is the joint probability of surviving to both T_1 and T_2 . This can be done by making use of the properties of Brownian Motion and the fact that the two events – the asset value at T_1 being greater than the endogenous default boundary and the asset value at T_2 being greater than the face value of the junior debt – are jointly lognormal. It can be shown that under the risk-neutral measure this is equal to $N_2(h_1^-, h_2^-; \rho)$ defined in the valuation equations above (see, Eq. 56.15). Thus, the total survival probability is:

$$\Pr\{(\ln V_{T_1} > \ln \bar{V}) \& (\ln V_{T_2} > \ln F_2)\} = N_2(h_1^-, h_2^-; \rho). \quad (23)$$

The forward default probability is easy to compute once the short-term survival probability and the total survival probability are known. The forward default probability is defined as the probability that the long-term debt obligations will not be satisfied given that the short-term debt obligations have been satisfied, or $\Pr\{V_{T_2} > F_2 | V_{T_1} > \bar{V}\}$. This is equal to one minus the conditional survival probability, or $1 - \Pr\{V_{T_2} > F_2 | V_{T_1} > \bar{V}\}$. The conditional survival probability is obtained by applying Bayes' Theorem to the short-term and total survival probabilities, which have already been proved to be $N(h_1^-)$ and $N_2(h_1^-, h_2^-; \rho)$, respectively. It can be shown that the conditional survival probability is $\frac{N_2(h_1^-, h_2^-; \rho)}{N(h_1^-)}$. Therefore the forward default probability is:

$$\Pr\{V_{T_2} \leq F_2 | V_{T_1} > \bar{V}\} = 1 - \frac{N_2(h_1^-, h_2^-; \rho)}{N(h_1^-)} \quad (24)$$

When there are n cash flows, there is a complete n -dimensional term structure of default probabilities (as well as survival probabilities) and $(n-1)$ -dimensional forward default probabilities. However, since the solutions are not known in closed-form everything must be computed using numerical methods (see, e.g., Chen et al. 2012).

56.4.2 Leland/Leland and Toft (LT)

Leland (1994) developed a structural credit risk model with endogenous default and, by including taxes and bankruptcy costs, was able to provide an analytical link between a firm's credit risk and its optimal capital structure. The model is set up as a first-passage problem and begins with many of the same assumptions as Black and Cox (1976). The asset value, V_t , evolves according to a Geometric Brownian Motion. Debt and equity are contingent claims on asset value, denoted by $D(V_t)$ and $E(V_t)$ respectively, both of which are known in closed form. The model assumes one class of debt with infinite maturity and a fixed coupon. The result is a flat, time-invariant default barrier, which is *endogenously determined* as the lowest possible asset value for which equity is still positive. Leland (1994) includes parameters for the corporate tax rate, τ , and bankruptcy cost, $\alpha \in [0, 1]$. This permits an optimal capital structure; something that previously had not appeared in a structural credit risk model (structural credit risk models typically assume that the capital structure decisions are exogenous to the model). To recap, in the Leland (1994) model the solutions for the values of debt and equity as well as the endogenous default barrier and the optimal capital structure are all known and given in closed-form.

Leland and Toft (1996) extend the Leland (1994) model to take debt maturity into account. All of the features from the original Leland (1994) model are preserved – including endogenous default, optimal capital structure, risky debt valuation, and complete closed-form solutions – with additional insights about the term structure of credit spreads and default probabilities. The Leland and Toft (1996) model assumes that every instant a fraction of the outstanding debt is retired at par and refinanced by issuing new debt with the same features (maturity, coupon, etc.) at the current market value. Essentially the firm is being modeled to systematically roll-over debt, which gives rise to a flat default barrier and a stationary debt structure.

The endogenous default barrier is derived using the “smooth-pasting condition.” The mathematics translates perfectly into economic intuition in justifying the default conditions. Since the shareholders look to maximize equity value, they will choose an optimal default boundary, V_B , that satisfies the following:

$$\left. \frac{\partial E(V; V_B, T)}{\partial V} \right|_{V=V_B} = 0 \quad (25)$$

The solution for the endogenous default boundary is:

$$V_B = \frac{\left(\frac{C}{r}\right)\left(\frac{A}{rT} - B\right) - A\left(\frac{P}{rT}\right) - \tau\left(\frac{C}{r}\right)(a + z)}{1 + \alpha(a + z) - (1 - \alpha)B} \quad (26)$$

where $a = \frac{(r - \delta - \frac{\sigma^2}{2})}{\sigma^2}$ and $z = \frac{[(a\sigma^2)^2 + 2r\sigma^2]^{1/2}}{\sigma^2}$;

$$\begin{aligned} A &= 2ae^{-rT}N(a\sigma\sqrt{T}) - 2zN(z\sigma\sqrt{T}) - \frac{2}{\sigma\sqrt{T}}N'(z\alpha\sqrt{T}) + \frac{2e^{-rT}}{\sigma\sqrt{T}}N'(a\sigma\sqrt{T}) \\ &\quad + (z - a) \\ B &= -\left(2z + \frac{2}{z\sigma^2T}\right)N(z\sigma\sqrt{T}) - \frac{2}{\sigma\sqrt{T}}N'(z\sigma\sqrt{T}) + (z - a) + \frac{1}{z\sigma^2T} \end{aligned}$$

$N(\cdot)$ is the cumulative standard normal distribution, i.e. $N(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$, and

$N'(\cdot)$ is the standard normal density function, i.e. $N'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Note that Eq. 26 is a function of the following model parameters:

- r : the risk-free interest rate
- σ : the asset volatility
- δ : the continuous payout rate
- α : bankruptcy cost (the asset write-down on default)
- τ : the corporate tax rate
- P : the total debt principal

- C : the total coupon payments to bondholders (per annum)
- T : the time-to-maturity of the debt.

This last point is very important. Whereas structural credit risk models with exogenous default specify the barrier independent of debt maturity, the endogenous default barrier of Leland and Toft (1996) is an explicit function of the debt maturity. This is what allows the authors to derive optimal maturity structure in addition to everything else. There is also a lot of useful insight that can be obtained by examining the comparative statics with respect to the endogenous default boundary. For instance, one can use Ito's Lemma to compute the total differential of equity as a function of asset value at the default boundary. Specifically,

$$dE|_{V=V_B} = \frac{1}{2}\sigma^2 V^2 \frac{\partial^2 E}{\partial V^2} \Big|_{V=V_B} dt. \quad (27)$$

This will illustrate precisely what happens to shareholders' wealth as asset value approaches the endogenous bankruptcy trigger. Eq. 27 simplifies to:

$$dE|_{V=V_B} = \left((1 - \tau)C + \frac{P}{T} \right) dt - (dD(V_B) + \delta V_B) dt \quad (28)$$

$$dE|_{V=V_B} = \left((1 - \tau)C + \frac{P}{T} - \frac{(1 - \alpha)}{T} V_B - \delta V_B \right) dt. \quad (29)$$

The first term in the parentheses in Eq. 29 is the annual after-tax coupon payment to bondholders; the second term is the annual rate at which debt principal is paid down; the third term can be shown to be the market value of debt at; and finally the last term is portion of asset value that is paid out over time dt . Letting $D(V)$ denote the market value of debt, then the market value of debt rolled over each year can be written as $\frac{D(V)}{T}$ and Eq. 29 can be rewritten as

$$dE|_{V=V_B} = \left[(1 - \tau)C + \frac{P}{T} \right] dt - \left[\frac{D(V_B)}{T} + \delta V_B \right] dt \quad (30)$$

The sum of the terms in the first set of brackets essentially shows the funds paid out annually: the after-tax coupon plus the face value of the debt that is being retired. The sum of the terms in the second set of brackets represents the funds that are coming in: the market value of new debt that is being raised (to replace the old debt that has been paid down) plus the cash-flow generated by the firm's assets. This is all happening in continuous time and therefore instantaneously. The left-hand-side of Eq. 30 is the incremental change in equity value at default as a function of asset value. So, at any given time, the firm is paying out the

[instantaneous] after-tax coupon plus the portion of debt that is being retired; simultaneously, funds are coming into the firm from both the issuance of new debt and the cash-flow generated by the firm's assets. If the instantaneous debt obligations (principal amount plus coupon payments) cannot be covered by the funds coming in, then shareholders must contribute the cash needed to fill the shortfall. The firm will default when the change in equity value is exactly equal to the additional amount needed to service the debt. Therefore the endogenous default condition implies that firm survival relies on the ability to raise capital. Note that this is very similar to the default condition in the compound option structural credit risk model discussed in the previous section. However, in the compound option structural model, recall that payments to debt holders were entirely financed by issuing new equity and the firm will default when the value of equity just equals the next cash flow. This condition ensures that, when the firm is solvent, new equity can be raised to pay down the maturing debts. This is in contrast to the Leland and Toft (1996) model that essentially requires raising equity only when the new debt raised cannot cover the maturing debt. Once the shareholders no longer find it advantageous to fill this gap, the firm will default. Since the debts are continuously being rolled-over the default boundary is constant resulting in a flat default barrier.

In the LT model, the probability of default can be computed analytically in terms of the first passage density using the endogenous default boundary, V_B , as the barrier. Define the first passage time (or the first hitting time) as

$$t^* = \inf\{t \geq 0 : V_t \geq V_B\} \quad (31)$$

then the default probability, or the probability of hitting V_B at or before time t^* , is

$$F(t^*) = N\left(\frac{-\ln\left(\frac{V}{V_B}\right) - \left(r - \delta - \frac{\sigma^2}{2}\right)t^*}{\sigma\sqrt{t^*}}\right) + \left(\frac{V}{V_B}\right)^{-2\left(\frac{r - \delta - \frac{\sigma^2}{2}}{\sigma^2}\right)} N\left(\frac{-\ln\left(\frac{V}{V_B}\right) + \left(r - \delta - \frac{\sigma^2}{2}\right)t^*}{\sigma\sqrt{t^*}}\right). \quad (32)$$

Leland (2004) examines the ability of structural credit risk models to fit historical default frequencies. Using aggregate-level data, he computes default probabilities using two barrier structural models (the endogenous LT model and the exogenous LS model) as well as the KMV Distance-to-Default method. Not surprisingly, all three models result in different default probability curves. The LT model fits the data fairly well for investment grade debt (Baa-rated), but underestimates default rates on speculative grade debt (single B rated). The LS model does not fit actual default frequencies well, except when the parameters are chosen so that it coincides with the endogenous barrier. The KMV model generates default probabilities that level off

over longer-term horizons, which is not observed in the empirical data.¹⁰ Furthermore, Leland (2004) shows that, contrary to popular belief, the KMV Distance-to-Default is not a sufficient statistic for probability of default.

56.5 Conclusion

This chapter surveyed the literature on structural credit risk models, specifically making the distinction between models with endogenous default versus those with exogenous default. Models with exogenous default impose an arbitrary default barrier, and therefore default probabilities can be manipulated by modifying the assumptions related the barrier. Endogenous models, on the other hand, do not assume arbitrary default conditions but rather solve for the point at which the firm will default under optimal conditions. Furthermore, they provide more insight in terms of the default process (when firms default) and allow for meaningful comparative statics (how default probabilities change with respect to the underlying variables). However, it was shown that imposing exogenous default allows for interesting real-world features to be incorporated into the model.

For instance, Longstaff and Schwartz (1995) and Collin-Dufresne and Goldstein (2001) both have structural credit risk models with exogenous default barriers that incorporate stochastic interest rates.¹¹ Longstaff and Schwartz (1995) explicitly allows for violations of absolute priority; the absolute priority rule, or APR, is quite often an assumption in most structural credit risk models. Their model was one of the first structural credit risk models that enabled the pricing of both fixed rate and floating rate debt. The implementation uses a relative valuation approach across a firm's debt classes and, in that respect, could potentially be applied to studying bond arbitrage.¹² Collin-Dufresne and Goldstein (2001) extended this model to allow for mean-reverting leverage ratios. Mean-reverting leverage ratios have been observed empirically and imply that firms issue and redeem debt to meet target capital structures. By incorporating this into a structural credit risk model they are able to generate credit spreads that are more consistent with those observed in financial markets.

The Geske model has endogenous default and made several major contributions. It was one of the first structural credit models to define default endogenously as the point at which a firm will no longer be able to raise capital to meet its forthcoming

¹⁰Typically the KMV model is used to estimate 1-year default probabilities, since it follows a European option setup. However, it is not uncommon for researchers to synthetically construct an ad-hoc default probability curve so as to allow for direct comparison between the KMV model and the barrier and compound option structural credit risk models which do give default probabilities over time.

¹¹Acharya and Carpenter (2002), not discussed here, have a structural credit risk model with endogenous default and stochastic interest rates.

¹²Schaefer and Strebulaev (2006) study capital structure arbitrage within the context of structural credit risk models. This is also discussed in Schaefer and Strebulaev (2008).

debt obligations. It was also the first structural credit risk model that allowed for coupon-paying debt, multiple classes of debt, and default to occur before maturity. The fact that, in the Geske model, default occurs at discrete points in time can be viewed as both a positive and a negative: it is actually realistic since defaults do typically occur when a major payment is due and not made; but with a large number of cash flows this becomes computationally inefficient. Furthermore, the fact that in the Geske model the firm is systematically de-leveraging is typically viewed as a negative feature. Essentially, this assumption implies that if the firm never defaults it will eventually have an all-equity capital structure.

The Leland (1994) model and its successor the Leland and Toft (1996) model represents the standard up to which all modern structural credit risk models are now held. The Leland and Toft (1996) model has a flat, endogenously determined default barrier. Also, since the model includes corporate taxes and bankruptcy costs it jointly values corporate securities and solves for the optimal capital structure. It also gives default probabilities for varying debt maturities. Leland (2004) compares default probabilities generated by the Longstaff and Schwartz (1995) exogenous barrier model with those generated by the Leland and Toft (1996) endogenous barrier model, and finds that the Longstaff-Schwartz model does not give reasonable default probabilities or recovery rates except when the parameters are chosen so that the default barrier coincides with the endogenous barrier solved by the Leland-Toft model.

References

- Acharya, V., and J.N. Carpenter. 2002. Corporate bond valuation and hedging with stochastic interest rates and endogenous bankruptcy. *Review of Financial Studies* 15 (5): 1355–1383.
- Bharath, S.T., and T. Shumway. 2008. Forecasting default with the merton distance to default model. *Review of Financial Studies* 21 (3): 1339–1369.
- Black, F., and J.C. Cox. 1976. Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance* 31 (2): 351–367.
- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81 (3): 637–654.
- Chen, R.-R., N. K. Chidambaran, M. B. Imerman, and B. J. Sopranzetti. 2012. Anatomy of financial institutions in crisis: Endogenous modeling of bank default risk. *Working Paper*.
- Collin-Dufresne, P., and R.S. Goldstein. 2001. Do credit spreads reflect stationary leverage ratios? *Journal of Finance* 56 (5): 1929–1957.
- Crosbie, P., and J. Bohn. 2003. *Modeling default risk*. Moody's KMV.
- Eom, Y.H., J. Helwege, and J.Z. Huang. 2004. Structural models of corporate bond pricing: An empirical analysis. *Review of Financial Studies* 17 (2): 499–544.
- Flannery, M.J., S. Nikolova, and O. Oztekin. 2011. Leverage expectations and bond credit spreads. *Journal of Financial and Quantitative Analysis*. forthcoming.
- Geske, R. 1977. The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis* 12 (4): 541–552.
- . 1979. The valuation of compound options. *Journal of Financial Economics* 7 (1): 63–81.
- Geske, R., and H.E. Johnson. 1984. The valuation of corporate liabilities as compound options: A correction. *Journal of Financial and Quantitative Analysis* 19 (2): 231–232.
- Goldstein, R., N. Ju, and H. Leland. 2001. An EBIT-based model of dynamic capital structure. *Journal of Business* 74 (4): 483–512.

- Jones, E.P., S.P. Mason, and E. Rosenfeld. 1984. Contingent claims analysis of corporate capital structures: An empirical investigation. *Journal of Finance* 39 (3): 611–625.
- Leland, H.E. 1994. Corporate debt value, BOND covenants, and optimal capital structure. *Journal of Finance* 49 (4): 1213–1252.
- . 2004. Predictions of default probabilities in structural models of debt. *Journal of Investment Management* 2 (2).
- Leland, H.E., and K.B. Toft. 1996. Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *Journal of Finance* 51 (3): 987–1019.
- Longstaff, F. A., and E. S. Schwartz. 1995. “A Simple approach to valuing risky fixed and floating rate debt,” *Journal of Finance*, 50(3), 789–819., Papers and Proceedings Fifty-Fifth Annual Meeting, American Finance Association, Washington, D.C., January 6–8, 1995.
- Merton, R.C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29 (2): 449–470.
- Schaefer, S., and I. Strebulaev. 2006. Risk in capital structure arbitrage. *Working Paper*.
- Schaefer, S., and I.A. Strebulaev. 2008. Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds. *Journal of Financial Economics* 90 (1): 1–19.
- Vasicek, O. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5: 177–188.



Arbitrage Opportunity Set and the Role of Corporations

57

James S. Ang and Yingmei Cheng

Abstract

We show that capital asset markets are made more efficient when investors and corporations could both engage in arbitrage. Previously known limits to arbitrage by investors are mitigated by explicitly recognizing the role of corporations as arbitrageurs. We derive the equilibrium regimes in which the dominant arbitrageurs in the capital markets are the investors, the corporations, both, or neither. Corporations enlarge the arbitrage opportunity set as they have several unique properties that could not be replicated by investors as arbitrageurs. We suggest that the desire to exploit arbitrage opportunities infeasible to investors could provide a new explanation for the formation of corporations.

Keywords

Corporate arbitrage · Investor arbitrage · Limits to arbitrage

57.1 Introduction

Arbitrage serves a critical role in the functioning of efficient capital markets. Viable arbitrage keeps the market prices of securities close to their fundamental values, and allows speedy asset price adjustment to new information. However, there are well-documented costs and risks to investors engaging in arbitrage. They range from information and transaction costs (Merton 1987) to fundamental risk and noise trader risk (Campbell and Kyle 1993; Shleifer and Vishny 1997). Limits to arbitrage by

J. S. Ang (✉)

Department of Finance, College of Business, Florida State University, Tallahassee, FL, USA
e-mail: jang@cob.fsu.edu

Y. Cheng

Department of Finance, Florida State University, Tallahassee, FL, USA
e-mail: ycheng@cob.fsu.edu

investors may prevent them from engaging in some arbitrage opportunities. If there were no alternatives to arbitrage by investors, market prices for some securities may deviate from their fundamental values for a sustained period (Shleifer 2000).

In this chapter, we demonstrate that corporations serve a useful function by stepping in to fill the role of arbitrageurs in many instances that the investors find infeasible or unprofitable. They can do so because the corporate form affords them certain advantages over investors in conducting arbitrages. Corporations can purchase mispriced assets in bulk and take possession of the control rights. These are essential to exploit arbitrage opportunities that are not available to the investors. Bulk purchase is required when the opportunity involves privately held businesses. Moreover, having control right enables the corporation to restructure (divide, repack-age, and sell) all or some of the assets of both private and public companies. Another principal advantage lies in the corporations' ability to create the short side of the arbitrage transaction by issuing and thus, shorting, their own securities. For instance, having identified an undervalued firm, an acquirer, as a corporate arbitrageur, can effectively construct an arbitrage position by acquiring the undervalued firm, and using its own stocks in exchange, i.e., shorting its own stocks. In contrast, the investors need to secure the short securities from third parties, incurring search and carrying costs. These differences suggest that corporations should be more capable of arbitraging long horizon mispricing, a shortcoming of investor arbitrage as pointed out by Shleifer and Vishny (1990).

Our analysis proceeds at two levels. We first model the technical aspects of putting together an arbitrage position by investors versus by corporations, such as carrying costs, purchase premium, fees and transaction costs, and the risk of early liquidation. The model yields an equilibrium framework delineating regimes in which dominant arbitrageurs responsible for market efficiency are identified. In the second part of the analysis, we show that, due to certain properties of the corporate form, there are types of market inefficiency that only corporations can exploit. Corporations expand the feasible corporate opportunity set, and thus, reduce the size and duration of inefficient asset pricing. Arbitrage by corporations complements arbitrage by investors. Investors may arbitrage away mispriced securities that involve relatively low transaction costs and short investment horizon. Corporations, on the hand, can do the 'heavy lifting': they engage in lumpier and longer term bets that investors avoid.

The contributions of this chapter are threefold. The first is to show that corporate arbitrage could be instrumental in keeping asset prices within tighter bounds of their long-term values. It helps us to understand better how capital markets achieve long-term efficiency. The second is to provide another rationale for the formation of corporations. Investors may get together to form corporations in order to conduct profitable arbitrages, since corporate arbitrage can be more feasible than the investor arbitrage under certain circumstances. The third is to shed new light on a good number of corporate policies and practices by interpreting them as attempts to conduct corporate arbitrage on mispriced assets. The examples range from acquisitions (Gonzales et al. 1998; Mandelbaum 1995; DeBondt and Thompson 1992), to restructuring involving major organizational changes such as bust up, buyout and spin off, to the timing and selection of different types of financing, and to the demand for financial innovations.

The remainder of the chapter is organized as follows. Section 57.2 reviews the literature on the limits to arbitrage. To compare investor and corporate arbitrage, we model, in Sect. 57.3, investors as arbitrageurs versus as shareholders of corporations engaging in arbitrage. We show that even under similar conditions, there exist a set of parameter values (circumstances) in which corporate arbitrage is profitable while investor arbitrage is not. Section 57.4 analyzes the types of arbitrage opportunities that are more likely to be exploited by corporations than by individuals. Section 57.5 discusses the implications of corporate arbitrage in the practice of corporate finance. Section 57.6 summarizes and concludes.

57.2 Limits to Investor Arbitrage

The classic definition of arbitrage requires that the arbitrageur be able to identify a mispriced security and its perfectly correlated counterpart, and take a long or short position in the mispriced security and an opposite position in its counterpart at zero net cost. These ideal conditions are generally conceded not to be attainable. There are costs and risks ranging from information and transaction costs (Merton 1987) to fundamental risk and noise trader risk (Campbell and Kyle 1993; Shleifer and Vishny 1997).

There are a number of papers that focus on the financial constraints faced by investor arbitrageurs. In Tuckman and Vila (1992), financial constraints can arise from holding costs. In Dow and Gorton (1994), financial constraints take the form of short horizon and trading costs. Shleifer and Vishny (1997) emphasize a wealth constraint in which money managers' arbitrage operation depends on the availability of external funds whose inflow is sensitive to the managers' performance. Gromb and Vayanos (2002) model financial constraints to arbitrageurs in the form of liquidation risks.¹ DeLong, Shleifer, Summers and Waldmann (1990) attribute limited arbitrage to noise trader risks and short horizons of risk-averse arbitrageurs. In Shleifer and Vishny (1992), investor arbitrageurs have short horizon, because the risks of adverse price movements and forced liquidation increase with holding period. Barberis and Thaler (2002) caution that there may not be a close substitute for the mispriced security. Wurgler and Zhuravskaya (2002) find that for stocks without close substitutes, arbitrage is weaker and mispricing is more frequent and severe. There is also the synchronization risk among arbitrageurs to affect price adjustments (Abreu and Brunnermeier 2002).

Short selling can be restrictive and costly (for example, D'Avolio 2002; Jones and Lamont 2002; Lamont and Thaler 2003). First, institutional restrictions and cultural bias limit short selling, for instance, Almazan, Brown, Carlson and Chapman (2004) document that although 30% of mutual funds are allowed to short, only 2% actually

¹For example, collateral value in a margin account becomes negative.

do. Lamont (2003) also finds that firms take actions to impede short selling. Second, the search for a stock lender could be costly and time consuming (Duffie et al. 2002). Third, short sellers are required to post additional collateral if prices rise (Mitchell et al. 2002). Fourth, stock loans are subject to recall risks, as the stocks that are being shorted may have to be returned to the lenders on demand. Short sellers may have to close their position and suffer a loss if they cannot find another lender. Finally, the security lending market can be illiquid (Ofek et al. 2004). The practical limits to short selling are reflected in the paucity of the number and frequency of actual short sales. Dechow, Hutton, and Meulbroek (2001) find evidence of few short sales: in 1976–1993, among stocks in NYSE (AMEX), 80% (90%) have less than 0.5% (5%) short to shares outstanding. In an international study, Bris, Goetzmann, and Zhu (2007) find markets that are more constrained for short sales have higher volatility and less informational efficiency.

To summarize, we find a burgeoning theoretical and empirical research in the economics and finance literatures recognizing the limits to arbitrage by investors. There are, however, still several important questions that have to be addressed: Are there alternatives to arbitrage by investors? When are corporations better at exploiting mispricing in the capital markets? What about stocks that investor arbitrageurs avoid: will their share prices be inefficient for prolonged periods, once they deviate from equilibrium? Is there a pricing bound to long term mispricing? In Sect. 57.4, we shall present arguments that corporations could fill the role of arbitrageurs in situations investors find infeasible. Corporate arbitrage would expand the set of arbitrage opportunities, and reduce the range of price inefficiency among long horizon assets. We compare arbitrage by investors versus by corporations at two levels: investors as arbitrageurs against shareholders of the corporations engaging in arbitrage (in Sect. 57.3), and against the corporations conducting arbitrage (in Sect. 57.4).

57.3 The Model

In the analysis to follow, we model the risk that prevents investors from committing to long horizon arbitrage: adverse price movement caused by noise traders that may result in premature liquidation. Since the risk of premature liquidation to investor arbitrageurs is related to their levered position and the associated risk of margin call in the interim periods, we endow investor arbitrageurs and shareholders of corporations engaging in arbitrage with the same amount of margined debt, to enable equitable comparison. The model below derives the arbitrage opportunity set for these two types of arbitrageurs. The size of the arbitrage opportunity set, being a function of the minimum mispricing for the arbitrage to be feasible, also gives a measure of the bounds on market inefficiency.

Here, we choose to model the case when the arbitrage opportunity is an undervalued asset/firm. It provides for potentially more interesting analysis as it involves more contending parties, with transactions that are harder to be replicated

by the investors. The analysis of the opposite case of arbitrage opportunities involving overvalued assets is straightforward (see footnote below).²

In the model, there are three dates, 0, 1 and 2. Date 0 is the starting time, when the noise trader may cause the stock of the firm to be mispriced, thus presenting an arbitrage opportunity. Date 1 is the intermediate period in which noise traders may cause the stock price to either diverge further away, or get closer to its fair value. Date 2 is the final time period in which asset price converges to its fair price, i.e., all uncertainties about pricing are resolved on this date.

Although the basic setting simplifies a multi-period model, the period between date 1 and date 2 may be long in practice, for example, several years. For the sake of simplicity, every agent in the model is assumed to be risk neutral.

57.3.1 Investor Arbitrageurs

In this basic model of investor arbitrageurs who face noise trader risks, we follow Shleifer and Vishny (1990) and Campbell and Kyle (1993) closely. The set of agents are: the investor arbitrageurs, the noise traders, and a firm with traded shares. Let V_t be the firm's fair value, and P_0 be its stock price at period 0. We assume the supply of shares is one unit. For simplicity, we further assume the mispriced firm's capital structure is made up of 100% equity.

At date 0, noise traders are pessimistic about the value of the firm, where α is the dollar size of the pessimistic shock. The quantity of the firm's equity demanded by the noise traders can be written as:

$$q(\text{noise}) = (V_t - \alpha)/P_0 \quad (1)$$

To the potential arbitrageurs, undervaluation of magnitude α gives rise to an arbitrage opportunity. To implement the arbitrage strategy, the investor arbitrageurs take a long dollar position of B in the stock of the undervalued firm at time 0. B is endogenously determined by the investor's maximization problem, as derived below. The quantity of the undervalued firm's equity demanded by the arbitrageurs is:

²The corporate arbitrageurs can arrange for a spinoff of their overvalued assets as independent legal entities, through IPOs to outside investors. The best-known example of exploiting an overvalued corporate asset is when 3Com spun off the overvalued Palm Pilot resulting in a valuation of Palm Pilot greater than that of its majority parent. They can also split off the overvalued assets and distribute pro rata shares of the new entities to existing shareholders. Finally, they can simply sell the overvalued asset to a willing buyer. Although investors can also sell shares of overvalued stocks they own, because they must hold diversified portfolios, potential gain from a single or few securities is small. Furthermore, information cost to uncover mispricing out of a portfolio with many stocks may reduce if not eliminate potential profits. The remaining question of how a firm comes upon overvalued asset is discussed in Sect. 57.5.

$$q(\text{arbitrageurs}) = B/P_0 \quad (2)$$

Because of unit stock supply, we have

$$P_0 = V_t - \alpha + B \quad (3)$$

To finance the arbitrage position of B dollars, the arbitrageurs need to short a similar security of equal amount. Let us call the security on the short side, S . Security S is highly but not perfectly correlated with the undervalued firm's security. For instance, they may be related to each other by many fundamental factors, but may not be affected by noise traders in the same way.³ Let net implementation cost for shorting security S be $B((1 + R)^n - 1)$, where n is the number of periods. For simplicity, we assume the risk-free interest rate is zero, thus R reflects the implementation cost for the investors to short the security. At date 1, the noise traders mistakenly identifying some noise as a fundamental change and deliver another shock, γ . The shock is evenly distributed between $-(V_t - \alpha)$ and $+\alpha$. That is, at date 1, the value of the firm, $V_t - \alpha + \gamma$ falls between 0 and V_t . The shock, even an optimistic one, is assumed not to be large enough to bring the security's market price back to its fundamental value at date 1. The distribution function of γ is $f(\gamma) = \frac{1}{V_t}$. To simplify, we shall assume the short security, S , has no noise trader risk at date 1, or that γ is net of common price variations between the two securities.⁴ We model noise trader risk as follows: when the arbitrageurs' asset value falls below a critical percentage of the original level, a margin call will be issued,⁵ and the arbitrageur will have to liquidate prematurely at date 1. Let k denote the critical level, where $k \in (0, 1)$. Given the firm's value is $V_t - \alpha$, the shock at date 1, γ , has to cause a price decline large enough to trigger a margin call, i.e., γ is negative and has a magnitude greater than or equal to $(1-k)(V_t - \alpha)$. If a margin call is issued, the arbitrageurs liquidate prematurely, and receive $(V_t - \alpha + \gamma) * (B/P_0)$. Otherwise, the arbitrageurs need to take no action at date 1, and will reach date 2 to receive final payoff of $V_t * (B/P_0)$. The arbitrageurs choose the optimal position size B by solving the following problem at date 0:

³Otherwise, the arbitrage position is perfect, and there is no risk, including that of premature liquidation.

⁴An alternate assumption is to let B be a loan, which will not experience noise trader risk.

⁵Another interesting complication in the literature is to allow arbitrageurs to invest more funds in the mispriced asset if it shall further deviate from its fair value, i.e., when there is a negative price shock at date 1, there are cash reserves to meet margin call and to commit more investments. However, the risk of this trading strategy increases as the number of periods, N , increases, as discussed in Sect. 57.5.

$$\begin{aligned}
& \text{Max}_{B>0} \frac{1}{1+R} \int_{-(V_t-\alpha)}^{-(1-k)(V_t-\alpha)} \frac{B}{P_0} (V_t - \alpha + \gamma) f(\gamma) d\gamma \\
& + \frac{1}{(1+R)^2} \int_{-(1-k)(V_t-\alpha)}^{\alpha} \frac{B}{P_0} V_t f(\gamma) d\gamma - B \\
& P_0 = V_t - \alpha + B.
\end{aligned} \tag{4}$$

That is, $\text{Max}_{B>0} \frac{1}{1+R} \frac{B}{V_t-\alpha+B} \frac{1}{V_t} \left[\frac{1}{2} k^2 (V_t - \alpha)^2 \right]$.

The endogenous nature of B is made evident by an examination of the above expressions. Increasing stake in the arbitrage position B , would lead to higher current price, P_0 , and thus, higher cost basis and beyond a certain point, lower profit for the arbitrageurs. The optimization problem may be interpreted as finding the optimal level of arbitrage activity by balancing the positive effect of B on profit due to size of the position, against the negative effect of B on investors' cost basis, while taking into account of premature liquidation. Net liquidation loss to the investors at date 1 increases with B .

The first order condition is:

$$\frac{V_t - \alpha}{(V_t - \alpha + B)^2} \cdot W = 1, \tag{5}$$

where $W = \frac{1}{1+R} \frac{1}{2V_t} k^2 (V_t - \alpha)^2 + \frac{\alpha + (1-k)(V_t-\alpha)}{(1+R)^2}$.

Solving Eq. 5, we have

$$B^* = \begin{cases} \sqrt{(V_t - \alpha)W} - V_t + \alpha, & \text{if } \sqrt{(V_t - \alpha)W} - V_t + \alpha > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Proposition 1 *Arbitrage by investors is feasible if and only if $\alpha \geq \alpha_0$, where,*

$\alpha_0 \equiv$

$$\frac{-\left[(1+R)^2 + k - (1+R)k^2\right] + \sqrt{\left[(1+R)^2 + k - (1+R)k^2\right]^2 + 2k^2(1+R)\left(-\frac{1}{2}(1+R)k^2 - (1-k) + (1+R)^2\right)}}{(1+R)k^2} \cdot V_t$$

and α_0 is a positive number.

The proof is in [Appendix A](#).

The intuition behind Proposition 1 is straightforward: the minimum magnitude of mispricing for the arbitrage to be profitable has to be large enough to offset the cost of implementing the position, and the risk of adverse price change at date 1. Alternatively, one may interpret $\alpha - \alpha_0$ as the size of the arbitrage opportunity set to investors, and the minimum value of α , i.e., α_0 , gives the range of admissible market inefficiency, where there is no investor arbitrage activity to eliminate mispricing.

Proposition 2 *Arbitrage by investors on mispriced securities is less than complete, i.e., it will not be enough to completely eliminate the amount of mispricing at date 0.*

Proof It is straightforward to verify that $B^* < \alpha$, i.e., $V_0 = V_t - \alpha + B < V_t$ at date 0. The details are in [Appendix A](#).

Because $P_0 = V_t - \alpha + B$, we have

$$P_0^* = \sqrt{(V_t - \alpha)W} \quad (7)$$

Equation 7 gives the equilibrium price at date 0, after investor arbitrageurs have completed their trading position.

Corollary 1 *At the optimal arbitrage position B , the expected payoff to the investor arbitrageurs is $\Lambda = [\sqrt{V_t - \alpha} - \sqrt{w}]^2$.*

The proof is in [Appendix A](#).

57.3.2 Shareholders of Corporations Engaging in Arbitrage

In this section we analyze the same set of issues (risk of liquidation, size of arbitrage opportunity set, admissible market efficiency, arbitrage profit) to shareholders of corporations engaging in arbitrage. Let V_0 be the market value of an undervalued firm, T , at date 0. A corporation A may recognize that firm T is underpriced, and construct an arbitrage position by paying a non negative premium (ρ) to acquire (long) all the stocks at $V_0 + \rho$, with its own stocks (the short component). The process may be described by the following time line: the acquirer proposes a merger and offers premium (ρ) at date 0, and the merger is completed before date 1 if the target accepts the merger offer at date 0. For simplicity, let us assume the capital structures of both the acquirer and the target are made up of 100% equity.

We analyze corporate arbitrage under the same set of circumstances as in Sect. 57.3.1. At date 0, firm T is underpriced by the amount α , that is: $V_0 = V_t - \alpha$. Misvaluation is to be resolved at date 2. At date 1, the noise traders, responding to a noise that is mistaken as a fundamental change, could cause the share price of firm T to change. The noise parameter γ has the same distribution as modeled in Sect. 57.3.1. Corporate arbitrageurs, in the first half of the arbitrage transaction, may take advantage of the opportunity by making an offer to firm T , now designated as the target firm. The shareholders of the target firms have two choices: to accept or reject the offer. If the target does not accept the premium offered by the acquirer, the target shareholders may get the margin call at date 1 if $V_t - \alpha + \gamma$ is less than $k(V_t - \alpha)$. Otherwise, the target shareholders hold the stocks until date 2 and receive the value V_t , since the target firm's undervaluation of α is corrected at date 2, i.e., $(V_t - \alpha) + \alpha = V_t$. The target's management is fully aware of size and timing of its mispricing α . That is, the expected payoff of the shareholders of the potential target, if not acquired, is

$$\begin{aligned}
& \int_{-(V_t-\alpha)}^{-(1-k)(V_t-\alpha)} (V_t - \alpha + \gamma) f(\gamma) d\gamma + \int_{-(1-k)(V_t-\alpha)}^{\alpha} V f(\gamma) d\gamma \\
&= \frac{1}{2} k^2 \frac{(V_t - \alpha)^2}{V_t} + (1 - k) V_t + \alpha k
\end{aligned} \tag{8}$$

The first term in Eq. 8 is the payoff in case of early liquidation, and the second term is the payoff of no early liquidation. If the target shareholders choose instead to accept the premium offered by the acquirer, we allow the same shock to the target's stock to affect share price of the merged firm, thus, cause the share price of the merged firm to decline by the same amount, that is, the value of the combined firm at date 1 is $V_a + V_t - \alpha + \gamma$, where V_a is the acquirer's value.⁶ If $V_a + V_t - \alpha + \gamma < k(V_a + V_t - \alpha)$, then the shareholders of the merged firm, too, receive the margin call. At date 2, the value of the merged firm, V_c , is the combination of the target's value, the acquirer's value and the incremental value Ω . That is, $V_c = V_a + V_t + \Omega$. Similar to correction from mispricing, α , the post merger incremental value Ω , is not to be realized until date 2. Ω can be positive or negative.

The shock, γ , is evenly distributed between $-(V_t - \alpha)$ and $+\alpha$. If $k < \frac{V_a}{V_a + V_t - \alpha}$, $V_a + V_t - \alpha + \gamma$ will always be greater than $k(V_a + V_t - \alpha)$. In this case, the risk of early liquidation does not exist for the shareholders of the combined firm. To consider the liquidation risk in the following analysis, we assume $k \geq \frac{V_a}{V_a + V_t - \alpha}$. The quantity in the right hand side is a function of the relative size of the acquirer to the pre revaluation combined value of the merged firm. That is, the effect of noise risk depends on the ability of the merged firm to absorb the shock. The results and the main conclusions will only become stronger in the case of $k < \frac{V_a}{V_a + V_t - \alpha}$, the analysis is presented in [Appendix B](#).

The target shareholders' expected payoff, in a stock for stock exchange, from the acquisition, is⁷:

$$\begin{aligned}
& \int_{-(V_t-\alpha)}^{-(1-k)(V_a+V_t-\alpha)} \frac{V_t - \alpha + \rho}{V_a + V_t - \alpha} (V_a + V_t - \alpha + \gamma) f(\gamma) d\gamma \\
& + \int_{-(1-k)(V_a+V_t-\alpha)}^{\alpha} \frac{V_t - \alpha + \rho}{V_a + V_t - \alpha} (V_a + V_t + \Omega) f(\gamma) d\gamma
\end{aligned} \tag{9}$$

We can rewrite Eq. 9 as $\frac{V_t - \alpha + \rho}{V_t(V_a + V_t - \alpha)} \cdot C$,

⁶Implicit in this formulation is that noise is a perturbation disguised as pricing factor that is capable of affecting share prices of the merged firm. The other way to model noise is to attribute mispricing from irrational trading of the target's shares by the noise traders. The latter case works in favor of corporate arbitrage since noise traders, not being able to trade in the now acquired target's shares, would not have any impact.

⁷The following formulation assumes the market attributes no gain to the merger at announcement, and the post announcement prices form the basis for the exchange ratio.

where $C = \frac{1}{2}[(V_a + V_t)(1 - k) + \alpha k]^2 + \frac{1}{2}V_t^2 + V_a V_t + \Omega((V_a + V_t)(1 - k) + \alpha k)$

C is positive as long as

$$\Omega > -\frac{\frac{1}{2}[(V_a + V_t)(1 - k) + \alpha k]^2 + \frac{1}{2}V_t^2 + V_a V_t}{(V_a + V_t)(1 - k) + \alpha k}. \quad (10)$$

Let us assume for convenience that Eq. 10 holds for now, which we will see later that this constraint is not binding.

The target's shareholders will accept the offer if

$$\frac{V_t - \alpha + \rho}{V_t(V_a + V_t - \alpha)} \cdot C \geq \frac{1}{2}k^2 \frac{(V_t - \alpha)^2}{V_t} + (1 - k)V_t + \alpha k \quad (11)$$

i.e., the premium ρ has to satisfy the following:

$$\rho \geq \alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1 - k)V_t^2 + \alpha k V_t}{C} \cdot (V_a + V_t - \alpha) \quad (12)$$

Because C increases in Ω , the minimum premium demanded by the target decreases in Ω . That is, a rational target would tradeoff merger premium and share of post merger expected gain in a stock for stock exchange; the higher the post merger gain is, the lower is the premium demanded by the target. We restrict the premium to be nonnegative, therefore the target demands the premium as, $\rho \geq$

$\max\left(\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1 - k)V_t^2 + \alpha k V_t}{C}(V_a + V_t - \alpha), 0\right)$. That is, (1) if $\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1 - k)V_t^2 + \alpha k V_t}{C}(V_a + V_t - \alpha) \geq 0$, the target demands the premium $\rho \geq \alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1 - k)V_t^2 + \alpha k V_t}{C}(V_a + V_t - \alpha)$; (2) if $\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1 - k)V_t^2 + \alpha k V_t}{C}(V_a + V_t - \alpha) < 0$, the target demand the premium of

0. $\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1 - k)V_t^2 + \alpha k V_t}{C}(V_a + V_t - \alpha) \geq 0$ is equivalent to $\Omega \leq \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1 - k)V_t^2 + \alpha k V_t}{(V_a + V_t)(1 - k) + \alpha k} - \frac{\frac{1}{2}[(V_a + V_t)(1 - k) + \alpha k]^2 - \frac{1}{2}V_t^2 - V_a V_t}{(V_a + V_t)(1 - k) + \alpha k}$. Let us denote $\frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1 - k)V_t^2 + \alpha k V_t}{(V_a + V_t)(1 - k) + \alpha k} - \frac{\frac{1}{2}[(V_a + V_t)(1 - k) + \alpha k]^2 - \frac{1}{2}V_t^2 - V_a V_t}{(V_a + V_t)(1 - k) + \alpha k}$ as Ω_0 .

The acquirer shareholder's expected value, in a stock for stock exchange, at date 2 is:

$$\int_{-(V_t - \alpha)}^{-(1 - k)(V_a + V_t - \alpha)} \frac{V_a - \rho}{V_a + V_t - \alpha} (V_a + V_t - \alpha + \gamma) f(\gamma) d\gamma + \int_{-(1 - k)(V_a + V_t - \alpha)}^{\alpha} \frac{V_a - \rho}{V_a + V_t - \alpha} (V_a + V_t + \Omega) f(\gamma) d\gamma - \phi \quad (13)$$

We can rewrite Eq. 13 as $\frac{V_a - \rho}{V_t(V_a + V_t - \alpha)} \cdot C - \phi$. ϕ is the acquirer's merger cost excluding the premium, for example, investment banking fees. ϕ is a positive number.

The acquirer shareholder's expected value, without the acquisition, at date 2 is V_a . Therefore, the acquirer will make the arbitrage motivated acquisition if

$$\frac{V_a - \rho}{V_t(V_a + V_t - \alpha)} \cdot C - \phi \geq V_a \quad (14)$$

That is, $\rho \leq V_a - \frac{(V_a + \phi)V_t}{C}(V_a + V_t - \alpha)$.

Because the premium is restricted to be nonnegative, the acquirer offers the following premium schedule:

If $V_a - \frac{(V_a + \phi)V_t}{C}(V_a + V_t - \alpha) > 0$, then $\rho \in \left[0, V_a - \frac{(V_a + \phi)V_t}{C}(V_a + V_t - \alpha)\right]$;

If $V_a - \frac{(V_a + \phi)V_t}{C}(V_a + V_t - \alpha) = 0$, then $\rho = 0$.

Proposition 3 *The necessary condition for the merger to be feasible is that $V_a - \frac{(V_a + \phi)V_t}{C}(V_a + V_t - \alpha) \geq 0$. That is,*

$$\Omega \geq \frac{\frac{(V_a + \phi)V_t}{V_a}(V_a + V_t - \alpha) - \frac{1}{2}[(V_a + V_t)(1 - k) + \alpha k]^2 - \frac{1}{2}V_t^2 - V_a V_t}{(V_a + V_t)(1 - k) + \alpha k}. \quad (15)$$

Proof If $V_a - \frac{(V_a + \phi)V_t}{C}(V_a + V_t - \alpha) < 0$, then the premium the acquirer is willing to offer has to be below zero, which violates the nonnegative premium constraint. Thus the merger is not feasible in that case.

Let us denote $\frac{\frac{(V_a + \phi)V_t}{V_a}(V_a + V_t - \alpha) - \frac{1}{2}[(V_a + V_t)(1 - k) + \alpha k]^2 - \frac{1}{2}V_t^2 - V_a V_t}{(V_a + V_t)(1 - k) + \alpha k}$ as Ω_1 . Ω_1 is a decreasing function of α . That is, the higher is the magnitude of undervaluation, the lower is the minimum value of post merger gain that is necessary for the acquirer to participate in the merger. Equation 15 implies that the assumption in Eq. 10 is not binding, as we mentioned before.

Proposition 4 *Given $\Omega \geq \Omega_1$, if $\Omega \geq \Omega_0$, the merger is feasible. Given $\Omega \geq \Omega_1$, if $\Omega < \Omega_0$, the merger is still feasible as long as $\Omega \geq \Omega_2 \equiv \frac{-\frac{1}{2}V_a^2(1 - k)^2 - V_a V_t(1 - k)^2 - \alpha V_a k(1 - k) + \phi V_t}{(1 - k)(V_a + V_t) + \alpha k}$.*

The proof is in [Appendix A](#).

Proposition 5 *The merger premium, ρ_0 , is less than the magnitude of misvaluation, α , if $\Omega > \Omega_3$, where Ω_3 is a negative number and*

$$\Omega_3 = \frac{-(\frac{1}{2}\alpha V_t^2 + \frac{1}{2}V_a^2 V_t + \frac{1}{2}V_a V_t^2)(1-k)^2 - k\alpha^2 V_t(1-k) - \frac{1}{2}(V_a V_t^2 - V_a \alpha^2 k^2 + \alpha^3 k^2 + \alpha V_t^2)}{V_t((V_a + V_t)(1-k) + \alpha k)}.$$

The proof is in [Appendix A](#).

Proposition 5 shows that when acquisition with premium is feasible for corporate arbitrageurs, the acquirer needs not give up all of arbitrage profit stemming from target's undervaluation in the form of premium, as long as the post merger gain is above a certain level.

In the equilibrium, the acquirer pays the minimum premium demanded by the target, i.e.,

$$\rho_0 = \max \left(\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{C} (V_a + V_t - \alpha), 0 \right) \quad (16)$$

We can derive the arbitrage profit for the shareholders of corporate acquirer as:

$$\begin{aligned} & \frac{V_a - \rho}{V_t(V_a + V_t - \alpha)} \cdot C - \phi - V_a \\ &= \min \left(\frac{C - \frac{1}{2}k^2(V_t - \alpha)^2 - (1-k)V_t^2 - \alpha k V_t}{V_t} \right. \\ & \quad \left. - \phi - V_a, \frac{V_a}{V_t(V_a + V_t - \alpha)} \cdot C - \phi - V_a \right) \end{aligned} \quad (17)$$

Corollary 2 For any magnitude of undervaluation, α , the merger is feasible as long as $\Omega \geq \max(\Omega_1, \Omega_0)$; when $\alpha \geq \alpha_1$, the merger is also feasible as long as $\Omega \geq \frac{-\frac{1}{2}V_a^2(1-k)^2 - V_a V_t(1-k)^2 - \alpha V_t k(1-k) + \phi V_t}{(1-k)(V_a + V_t) + \alpha k}$, where

$$\alpha_1 = \frac{-[(V_a + \phi)V_t - k^2 V_a V_t] + \sqrt{[(V_a + \phi)V_t - k^2 V_a V_t]^2 + 2k^2 V_a (k V_a V_t^2 - \frac{1}{2}k^2 V_a V_t^2 + \phi V_t^2)}}{V_a k^2}.$$

The proof is in [Appendix A](#).

Proposition 4 and Corollary 2 demonstrate that regardless of the magnitude of misvaluation, α , the merger is feasible if the post merger gain, Ω , is above a certain level. Base on the results we obtained so far, we can specify the sets of parameters under which the investor and/or corporate arbitrage may be feasible.

Table 1 shows that there are four possible equilibria in different regimes: investors only, corporations only, both investors and corporations, and neither type.

Table 1 Equilibrium regimes of dominant arbitrageurs

	$\Omega < \Omega_1$	$\Omega \geq \Omega_1$, and $\Omega \geq \Omega_0$	$\Omega \geq \Omega_1$, and $\Omega < \Omega_0$
$\alpha < \alpha_0$, and $\alpha < \alpha_1$	Neither	Corporations	Not applicable ^a
$\alpha < \alpha_0$, and $\alpha > \alpha_1$	Neither	Corporations	Corporations if $\Omega \geq \Omega_2$
$\alpha > \alpha_0$, and $\alpha < \alpha_1$	Investors	Corporations, and investors	Not applicable ^a
$\alpha > \alpha_0$, and $\alpha > \alpha_1$	Investors	Corporations, and investors	Corporations if $\Omega \geq \Omega_2$; and investors

^a“ $\Omega \geq \Omega_1$, and $\Omega < \Omega_0$ ” requires that $\alpha > \alpha_1$

$$\alpha_0 \equiv \frac{-[(1+R)^2 + k - (1+R)k^2] + \sqrt{[(1+R)^2 + k - (1+R)k^2]^2 + 2k^2(1+R)(-\frac{1}{2}(1+R)k^2 - (1-k) + (1+R)^2)}}{(1+R)k^2} \cdot V_t$$

$$\text{and } \alpha_1 = \frac{-[(V_a + \phi)V_t - k^2 V_a V_t] + \sqrt{[(V_a + \phi)V_t - k^2 V_a V_t]^2 + 2k^2 \alpha^2 V_a (kV_a V_t^2 - \frac{1}{2}k^2 V_a V_t^2 + \phi V_t^2)}}{V_a k^2}$$

These regimes are classified on basis of two sets of parameters: the size of original pre arbitrage opportunity, and the size of post arbitrage (or merger) gains. In presence of large mispricing but low post merger gain, investor arbitrage is favored. But corporate arbitrage is feasible, regardless of target's mispricing, if post merger gain is higher than a certain level, e.g., $\Omega \geq \max(\Omega_1, \Omega_0)$. In regimes admitting both types of arbitrageurs, several scenarios may be possible: one is that the investor arbitrageurs would initiate arbitrage position to put the target in play, or to flush out potential acquirers; another is that the investor arbitrageurs engage in the arbitrage after an acquisition is announced but the probability of the merger to reach completion is highly uncertain.⁸ Finally, there exists a regime in which neither investors nor corporations could engage in arbitrage, and market may remain inefficient. In this regime, the pre-merger undervaluation is small and the post-merger gain is low.

Both the traditional and behavioral finance literatures recognize that limits to arbitrage for individuals and institutions may provide the reasons for well-documented instances of market failures, e.g., market anomalies and price bubbles. Since both risks and costs of an arbitrage position increase with time, limits to arbitrage are especially severe when the period to mispricing correction is long (Shleifer and Vishny 1990). The period of apparent inefficiency may increase with the time needed to reveal information.

⁸In the topic of risk arbitrage (see Cornelli and Li 2002; Mitchell and Pulvino 2002), investors and corporations may be regarded as arbitrageurs at the same time, with the common goal of seeing the merger completed.

57.4 Expanding the Arbitrage Opportunity Set by Corporations

Up to this point, we limit our analysis to arbitrage opportunities that are available to both investors and corporations, with emphasis on the technical aspects of putting together an arbitrage position by investors versus by corporations, such as carrying costs, purchase premium, availability and cost of the short position, and the risk of early liquidation. The analysis is necessary as it enable us to make comparison between these two alternatives under similar circumstances. However, the analysis of arbitrage by corporations is not complete since their arbitrage opportunity set is much larger, i.e., there are situations corporations could arbitrage but investors could not. We shall in this section analyze those corporate arbitrage opportunities, and discuss in the next section the implications of this expanded ability to arbitrage.

An importance difference between an investor and a corporation as an arbitrageur is that the corporation may be able to engage in ‘bulk’ purchase. With the acquisition of the entire target firm, come the ownership and management rights. These sets of rights are particularly important to unlock the arbitrage profits hidden under some assets. Examples that may involve significant arbitrage gains are provided in the following:

1. When privately held companies are mispriced, they are simply not available to investors.⁹ An examination of merger data between 1977 and 2001 (the data source is SDC) indicates that the ratio of the numbers of private targets to public targets is close to 3:2. Although not all such mergers were motivated by arbitrage, but for those that were, it is certain that investor arbitrageurs could not participate.
2. Corporations can purchase the assets in one piece, and create value by fission, i.e., selling parts that are mispriced (Fluck and Lynch 1999). For example, in “bust up” transactions, the corporate arbitrageurs acquire complete property rights before assets are sold or resold separately.¹⁰
3. The corporations can also create value by fusion, e.g., purchasing separate assets to form new enterprises. Examples include combining parts of the supply chain to form an integrated company, or acquiring several broadcasting stations to sell them as a national network.¹¹ Investors simply could not replicate these transactions.

⁹Such information on mispricing may be initiated from an analysis of comparable public companies, as private firms, or their consultants and bankers, as sellers often use valuation of comparable public companies as benchmark in pricing.

¹⁰Coyne, Fabozzi, and Yaari (1992) have shown how acquisitions may be conducted for asset arbitrage given various business and personal taxes.

¹¹There was the case of John Kluge who put together the Metromedia Broadcasting Company that was later sold to become the Fox Network, etc.

4. Corporate arbitrageurs take control of the misvalued firm, and can influence the timing and size of price corrections. While individual and institutions are by and large passive investors, corporations take an active role in the management of the assets they acquire. Once they acquire complete property right, they can restructure through change in top management, employee layoff, and other cost cutting measures. Furthermore, acquirers could supply the necessary information technology, production skills and facilities, human capital, marketing network, or relatively inexpensive financing for the targets to realize the added value, or synergy. These actions, if effective, would advance the date of price correction.
5. Very substantial information costs are required to identify some of the potential arbitrage opportunities. Corporations, as bulk purchasers, can economize information costs. The ability of the corporate acquirers to conduct “due process” enables them to obtain more accurate information than what the investor arbitrageurs could, and in the case of adverse information, the acquirers can unwind the position by withdrawing the merger offer. In addition, by purchasing the entire asset, they can avoid high transaction costs from forming and unwinding positions on securities with low liquidity.

The above lists the advantages of corporate arbitrageurs over investor arbitrageurs on the long side. Another set of advantages is associated with the short side of the trading position. Specifically, these are:

1. Firms can short their own stocks by exchanging their own securities for mispriced assets or securities. This is in contrast to investor arbitrageurs having to locate the short securities from third parties and to pay carrying costs. Investor arbitrageurs face a hard constraint when selling borrowed shares: the shares have to be returned at the close of arbitrage position if not sooner. Thus, they have to absorb all the residual risks that are not covered by the arbitrage position. Carrying costs are also mandatory. On the other hand, corporate arbitrageurs face a soft constraint when exchanging their own shares to acquire undervalued assets.¹² There is no legal or compelling requirement for the corporations to settle their positions at a definite date. Further, recipients of the corporate arbitrageurs’ securities, who were former target shareholders, have to share the residual risks. Carrying costs, such as dividends, are not mandatory. If both types of arbitrageurs choose to finance their positions with debt, corporate arbitrageurs can shield their shareholders with the limited liability provision, while individuals may risk personal bankruptcy, and institutions may have to contend with institutional or legal restrictions on the use of leverage. As the arbitrage horizon, N , increases, risks as well as other limits to investor arbitrage are magnified by a greater extent than

¹²We emphasize the word exchange to call attention to the difference between new shares issued to outsiders as in SEOs (secondary equity offers), and an exchange of shares in a merger.

those to corporate arbitrageurs. If noise trader risk could occur in each of the N periods, the overall risk of premature liquidation for investor arbitrageurs increases as N increases. As the number of time period increases, the overall chance that prices of the originally mispriced security may be further away before converging to its eventual fair value is higher.¹³ Furthermore, longer time periods would make unattractive, if not eliminate, the strategy of investing even more money when noise trader risk causes another adverse price response in the next period. As the number of periods increases, committing more funds in response to current losses approaches a form of gambler's fallacy, expecting the probability of a favorable reversal to rise. By contrast, the corporate arbitrageurs/acquirers, not facing the possibility of having to settle up in each period, can afford to be patient.¹⁴

2. Since firms may issue overvalued securities, corporate arbitrageurs have the unique opportunity to profit from both sides of the arbitrage position. Although investor arbitrageurs may long underpriced securities and short overpriced securities, there are numerous practical difficulties. First, without corporate arbitrageurs' inside informational advantage, investor arbitrageurs would be less likely to identify arbitrage opportunities with potential two-way gains. Second, lenders of short securities, who are long-term holders, may charge a higher carrying cost for overpriced securities. Third, investor arbitrageurs face greater risks, as there is no guarantee that price corrections for both securities would occur at the same time when the trading positions are to close, that is, there is a synchronization risk (Abreu and Brunnermeier 2002). In contrast, corporate arbitrageurs have more leeway with respect to lack of synchronization in price corrections, since the arbitrageurs do not have to settle at a definite date.

In addition to the synchronization risk mentioned above, investors also face: (1) the inconvenience of recall risk by "short" lenders, in which the replacement has to be found; (2) financing risk, where additional collateral may be needed and thus render the arbitrage no longer a zero investment trading position; (3) horizon risk, i.e., risk of loss from imperfect arbitrage increases with the length of the holding period; (4) fundamental risk, i.e., no arbitrageurs can be certain of the value of the securities; and (5) noise trader risk, as noise traders may cause mispriced securities to diverge further away from their fundamental values. Large blockholders may be able to economize some of the transaction and

¹³This source of risk increases even further if N is stochastic, thus creating uncertain termination date that could be further confounded with liquidity and synchronization risks.

¹⁴We acknowledge that certain combination of circumstances such as large cumulative negative price shocks, large target relative to the acquirer, and high correlation between noise trader risks on acquirer's and target's securities can render the combined firm an undervalued and attractive target to other firms. However, this risk should be smaller than the risk of premature liquidation by investor arbitrageurs.

searching costs, but they still have to find ways to obtain the securities to short, and bear the risks from imperfect arbitrage.¹⁵

Corporate arbitrageurs, who take the target's shares off the market and create their own short securities by issuing stocks, can avoid synchronization risk, recall risk, and noise trader risk altogether. Financing risk is greatly reduced if the asset acquired is highly correlated with existing assets, such as in horizontal acquisitions.¹⁶

Fundamental risk could be reduced as corporations have superior information about the targets. Arbitrage trading position is often not riskless, since there are uncertainties concerning the fundamental value of the security in question. The fair value of the security, when finally revealed, may deviate from the value expected at date 0, imposing unknown risks on the arbitrageurs if the short position does not perfectly correlate with the misvalued security. Although both investor arbitrageurs and corporate arbitrageurs are subject to the fundamental risks, the way each type deals with the risk could be different.

Investor arbitrageurs have to settle up at date 2, and suffer the entire loss. Shareholders of the acquirer, however, are able to share some of the loss with the target shareholders who are also stockholders of the same firm. This enables the original shareholders of the corporate arbitrageurs to absorb fundamental risks better than investor arbitrageurs.

Finally, corporations can better absorb horizon risks. They may have lower carrying or financing costs. They are more able to maintain a long holding period as they face less pressure to unwind, and they are more actively involved in the management of the assets to hasten the price correction.

Table 2 provides a summary of the main differences between investor arbitrage and corporate arbitrage. Because small investors, in comparison to large individual and institutional investors, are handicapped due to their greater information and transaction costs, the table compares corporate arbitrageurs with only the most efficient investor groups, i.e., large individual and institutional investors. Because of their advantages, firms are capable of expanding the scope of arbitrage opportunities in several ways.

To summarize, these extensions discussed above reinforce the importance of corporate arbitrage, especially when the period of misvaluation correction is long. Since active involvement by the acquiring firms could shorten the time to correct misvaluation, corporate arbitrage helps to facilitate the long-term market efficiency.

¹⁵For them to overcome the bulk purchase problem, they need to be able to acquire the decision right via control of the board. That will incur control contest costs, and they have to face free riding cost later on.

¹⁶It is of interest to note that when a corporate arbitrageur acquires a related company, the cash flows from the stocks of the combined firm offered (the short position) should match the cash flows to be generated from the target's assets. That is, the cash flows from the target's own assets support the stocks received by the target's shareholders. In other words, this type of arbitrage is *self-financing*, satisfying one of the requirements of an ideal arbitrage in which all financing risks to the acquirers should be eliminated.

Table 2 A comparison of large investors versus corporate arbitrageurs

	Large investors, including institutions	Corporate arbitrageurs
<i>1. Arbitrage opportunities involving bulk purchase</i>	Can only make fractional purchase of traded assets, preferably in liquid markets	Can make bulk purchase
(a) Private firms	No	Yes
(b) Bust up deals or fissions	No	Yes
(c) Create new business entities from parts, or fusions	No	Yes
<i>2. Identifying, executing and realizing arbitrage opportunities</i>		
(a) Information costs	Pay same fixed cost	Can spread cost over 100% of assets
(b) Costs of forming and unwinding illiquid assets	Pay premium to acquire meaningful holding; costs to liquidate could be high	Pay higher premium to acquire 100%; but exit price may be higher due to transfer of control rights
(c) Ability to affect size and timing of misvaluation correction	None as passive investors, or limited through pressure on management via governance channels	Yes, they can actively restructure operations; incorporate earnings improvement in own financial reports
<i>3. Ease in obtaining the securities to short</i>		
(a) Source of “short” securities	Have to borrow shares from willing lenders by incurring searching and carrying costs	Create “homemade short” by issuing own securities
(b) Extra opportunity to arbitrage	None	Possible, can issue own overvalued securities
<i>4. Dealing with imperfect arbitrage</i>		
(a) Synchronization risk	Forming and unwinding both the long and short position may not occur at the same time	More flexibility with resolution date, as own stocks have no finite maturity
(b) Recall risk	Yes, an inconvenience in normal period, but could be costly when needed most	None. “Short” investors hold corporation’s securities
(c) Financing risk	Collateral in the form of cash and/or securities makes investor arbitrage not a net zero investment strategy. Additional residual risk may be introduced from less than perfect correlation between long and short assets	Small or none, arbitrage is self-financing in all future periods if asset acquired with stocks is highly correlated with own assets
(d) Horizon risk	Risk of loss from imperfection in arbitrage increases with length of time horizon	Greater ability to absorb horizon risk due to a combination of several

(continued)

Table 2 (continued)

	Large investors, including institutions	Corporate arbitrageurs
		factors: lower financing costs, active involvement, homemade short, and less pressure to unwind and settle in finite date
(e) Fundamental risk	Could increase unhedged residual risk	May minimize valuation risks if acquisition of undervalued targets from industries where corporate arbitrageur has information advantage
(f) Noise trader risk	Could cause premature liquidation of arbitrage position. Large institutions with diversified portfolio may reduce but cannot eliminate liquidation risks	Noise trader could not affect prices of assets that are taken off the market by corporate acquirer

57.5 Implications of Corporate Arbitrage

Finally, we consider the implications of corporate arbitrage on the theories and applications of corporate finance. First, there is an implicit assumption in the Modigliani and Miller (M&M) theories: arbitrage by investors and by corporations are equivalent. The classic proof in M&M uses an argument that the individual investors are able to emulate the firm's balance sheet in the personal accounts. As pointed out above, and summarized in Table 2, there are arbitrage opportunities that could only be undertaken by the corporate entities. In particular, there are assets, lumpy or in private hands, that only corporations could acquire. In addition, corporations can issue new equities, thus changing their capital structure, to finance the type of arbitrage that investors find infeasible or unprofitable. M&M's emulation strategy breakdowns here as investors could neither acquire the same asset nor duplicate the same capital structure. Thus, when the twin M&M implicit assumptions of efficient capital market pricing, and perfect replication of corporate business and financial policies by investors, are relaxed, we have the following:

A Capital Structure Proposition:

When there arise arbitrage opportunities in which corporations but not investors can exploit, optimal capital structure changes in response to these opportunities.

Specifically, for the value maximizing firms, (1) arbitrage involving undervalued assets decreases leverage; (2) arbitrage involving overvalued assets may increase or have neutral effect on leverage. Leverage reduces when the firms issue stocks to finance purchase of undervalued assets. Leverage may increase when equity shares are distributed to shareholders in the new entity, or does not change when cash is

received from IPO of spin off or outright sale of the overvalued assets, in which cash could retire both debt and equity proportionately.

Since the arbitrage (by investors) mechanism is a cornerstone in the M&M derivations, viable corporate arbitrage, when investor arbitrage fails, is expected to affect their theories as much as the relaxation of their other assumptions, such as no taxes, transaction costs, distress costs, agency problems, and information asymmetry.

Second, if corporate arbitrage can be shown to be more feasible than investor arbitrage in at least some circumstances, the corporate organizational form may be justified from a purely financial standpoint. Rational investors pool their funds together to form corporations in order to exploit arbitrage opportunities that they could not have engaged in by themselves. A special case is a theory of conglomerates, and thus unrelated diversification, in a world in which there exists a positive probability of mispricing in future periods. Without assuming that the firms have the ability to predict which stocks may be mispriced in the future, conglomerate firms that started out in time 0 with a collection of unrelated assets know that some assets may be overvalued or undervalued in future periods. For the assets that turn out to be overvalued at a later date, the conglomerate now has overvalued assets to spinoff. And for the assets that turn out to be undervalued in the future, because mispriced assets are small in proportion to the conglomerate's size, the conglomerate structure shields undervalued assets from becoming the targets of acquisition, a fate that is more likely to befall stand alone undervalued firms.

Third, we can rethink various corporate events as instances of corporate arbitrage in situations when investor arbitrage is not possible. These include: overvalued firms acquiring undervalued firms with stocks; firms spinning off or issue as tracking stocks lines of business that are overvalued by the market at the moment; firms issuing overvalued stocks to buy back or retire less overvalued debt; firms issuing stocks to pay dividends and boost stock prices when high dividend paying value stocks are in high demand; firms borrowing at market prices to repurchase undervalued stocks; firms engaging in leverage buyouts by buying the firm's asset and shorting or borrowing outside funds when gross undervaluation occurs; private firms choosing between selling itself to another firm and issuing an IPO (depending on which market is more optimistic – this is an example of arbitrage between real and capital markets); firms engaging in interest rate and currency swaps (for instance, firms facing high long-term financing rate but low short-term rate can issue short and swap into long-term); and firms forming business alliances, partial ownership arrangements, and cross shareholdings to avoid diversification.

57.6 Conclusion

We start with the question: when are corporations better than investors in exploiting mispricing in the capital markets? The arbitrage opportunity sets for both investor and corporate arbitrage are studied in our model. Each arbitrage opportunity set

specifies the maximum bounds within which capital assets may not be efficient. Previously known limits to arbitrage by investors are mitigated by explicitly recognizing the role of corporations as arbitrageurs. Corporations have several desirable properties as arbitrageurs in dealing with lumpy investments, privately held concerns, those requiring control of property rights, mispricing requiring longer horizon to resolve, etc. The corporate arbitrageurs also have the ability to minimize the cost and risks of taking the short side and arbitrage.

The theoretical significance of corporations participating in arbitrage is that not only they expand the arbitrage opportunity set and thus, reducing the size of mispricing set, to help facilitate capital markets to be more efficient, but also in the way they influence the quality of market due to the type of assets they choose to arbitrage. With the limitations on the investors' ability to arbitrage, there could be serious deficiencies in the capital markets where large and persistent mispricing may occur for assets that would take time to correct, or effort via restructuring. Corporate arbitrageurs fill this gap, as they can engage in long horizon and lumpy arbitrage.

In our model, we demonstrate that investor arbitrage and corporate arbitrage can both exist. Viable corporate arbitrage complements investor arbitrage in fostering efficient capital market operations. Investor arbitrageurs are known to facilitate market prices toward short-term equilibrium, while we show here that corporate arbitrageurs facilitate toward long-term equilibrium. When noise trader risk or implementation cost is high, and the period of price correction is long, corporate arbitrage is more feasible. Corporate arbitrage provides a new insight into the function of firms. Firms are able to conduct arbitrage in situations where investor arbitrage is not feasible, thus providing a pure financial rationale for the existence of corporation. Many observed corporate financial practices can also be explained as ways to facilitate arbitrage on a large scale, from buying large firms in order to breakup later, to taking firms private, to timing and types of financing, etc.

Appendix A: Proofs of Propositions and Corollaries

A.1. Proof of Proposition 1

Proof $\alpha \geq V_t - \sqrt{(V_t - \alpha)W}$ is equivalent to $V_t - \alpha \leq W$. That is:

$$V_t - \alpha \leq \frac{1}{1+R} \frac{1}{2V_t} k^2 (V_t - \alpha)^2 + \frac{\alpha + (1-k)(V_t - \alpha)}{(1+R)^2}.$$

If we collect the terms in the above inequality, we have:

$$\begin{aligned} & \frac{1+R}{2V_t} k^2 \alpha^2 + \left[(1+R)^2 + k - (1+R)k^2 \right] \alpha + \frac{1+R}{2} k^2 V_t \\ & + (1-k)V_t - (1+R)^2 V_t \geq 0 \end{aligned}$$

$$\alpha \geq \frac{-[(1+R)^2 + k - (1+R)k^2] + \sqrt{[(1+R)^2 + k - (1+R)k^2]^2 + 2k^2(1+R)\left(-\frac{1}{2}(1+R)k^2 - (1-k) + (1+R)^2\right)}}{(1+R)k^2} \cdot V_t$$

$$\text{Let } \alpha_0 \equiv \frac{[(1+R)^2 + k - (1+R)k^2] + \sqrt{[(1+R)^2 + k - (1+R)k^2]^2 + 2k^2(1+R)\left(-\frac{1}{2}(1+R)k^2 - (1-k)(1+R)^2\right)}}{(1+R)k^2} \cdot V_t$$

Then α must satisfy the following:

$$-\frac{1}{2}(1+R)k^2 - (1-k) + (1+R)^2 = R^2 + 2R + k - \frac{1}{2}k^2 - \frac{1}{2}Rk^2 > 0, \text{ thus } \alpha_0 > 0.$$

A.2. Proof of Proposition 2

Proof

$$\begin{aligned} B - \alpha &= \sqrt{(V_t - \alpha)W} - V_t + \alpha - \alpha = \sqrt{(V_t - \alpha)W} - V_t \\ (V_t - \alpha)W &= (V_t - \alpha) \left\{ \frac{1}{1+R} \frac{1}{2V_t} k^2 (V_t - \alpha)^2 + \frac{\alpha + (1-k)(V_t - \alpha)}{(1+R)^2} \right\} \\ &< (V_t - \alpha) \left[\frac{1}{2V_t} k^2 (V_t - \alpha)^2 + \alpha + (1-k)(V_t - \alpha) \right] \\ &= \frac{V_t - \alpha}{2V_t} \left[k^2 (V_t - \alpha)^2 - 2V_t k (V_t - \alpha) + 2V_t^2 \right] \\ &= \frac{V_t - \alpha}{2V_t} \left[(k(V_t - \alpha) - V_t)^2 + V_t^2 \right] \\ &< \frac{V_t - \alpha}{2V_t} \cdot 2V_t^2 = (V_t - \alpha)V_t < V_t^2 \end{aligned}$$

That is, $\sqrt{(V_t - \alpha)W} < V_t$. Therefore, $B - \alpha < V_t - V_t = 0$.

A.3. Proof of Corollary 1

The arbitrageurs' payoff is:

$$\begin{aligned} \frac{WB}{V_t - \alpha + B} - B &= \frac{W \cdot (\sqrt{(V_t - \alpha)W} - V_t + \alpha)}{\sqrt{(V_t - \alpha)W}} - \sqrt{(V_t - \alpha)W} + (V_t - \alpha) \\ &= W - 2\sqrt{(V_t - \alpha)W} + (V_t - \alpha) \\ &= \left(\sqrt{V_t - \alpha} - \sqrt{W} \right)^2 \end{aligned}$$

A.4. Proof of Proposition 4

Proof If $\Omega \geq \Omega_0$, the minimum premium the target demands is zero. Given $\Omega \geq \Omega_1$, the premium that the acquirer is willing to pay is no less than zero. Therefore, the acquirer pays zero premium and the merger is feasible.

If $\Omega < \Omega_0$, the target demands a positive premium: $\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{C}(V_a + V_t - \alpha)$. The sufficient condition for the acquisition to be feasible is that the maximum premium the acquirer is willing to pay is no less than the minimum premium demanded by the target, that is:

$$V_a - \frac{(V_a + \phi)V_t}{C}(V_a + V_t - \alpha) > \alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{C}(V_a + V_t - \alpha)$$

It is equivalent to:

$$\begin{aligned} & C - \frac{1}{2}k^2(V_t - \alpha)^2 - (1-k)V_t^2 - \alpha k V_t - (V_a + \phi)V_t > 0 \\ & C - \frac{1}{2}k^2(V_t - \alpha)^2 - (1-k)V_t^2 - \alpha k V_t - (V_a + \phi)V_t \\ & \quad (V_a + V_t)(V_t + \Omega) - k(V_a + V_t)(V_a + V_t - \alpha + \Omega) \\ & = +k\alpha\Omega + \frac{1}{2}V_a^2 + V_a V_t + \frac{1}{2}k^2(V_a + V_t - \alpha)^2 - \frac{1}{2}k^2(V_t - \alpha)^2 \\ & \quad - (1-k)V_t^2 - \alpha k V_t - V_a V_t - \phi V_t \\ & V_a V_t + V_a \Omega + \Omega E_t - k V_a^2 - 2k V_a V_t + \alpha k V_a \\ & = -\Omega k V_a - \Omega k V_t + k\alpha\Omega + \frac{1}{2}V_a^2 + \frac{1}{2}k^2 V_a^2 + k^2 V_a V_t - \alpha k^2 V_a - \phi V_t \\ & = \frac{1}{2}V_a^2(1-k)^2 + V_a V_t(1-k)^2 + \Omega V_a(1-k) + \alpha k V_a(1-k) \\ & \quad + \Omega V_t(1-k) + \alpha\Omega k - \phi V_t \end{aligned}$$

Therefore Eq. 18 holds if and only if

$$\begin{aligned} & \Omega[(1-k)(V_a + V_t) + \alpha k] + \frac{1}{2}V_a^2(1-k)^2 + V_a V_t(1-k)^2 \\ & + \alpha V_a k(1-k) - \phi V_t > 0 \end{aligned}$$

That is:

$$\Omega > -\frac{\frac{1}{2}V_a^2(1-k)^2 + V_a V_t(1-k)^2 + \alpha V_a k(1-k) - \phi V_t}{(1-k)(V_a + V_t) + \alpha k} \quad (18)$$

A.5. Proof of Proposition 5

Proof

$$\begin{aligned}\rho_0 - \alpha &= \max \left(\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{C} \right. \\ &\quad \left. (V_a + V_t - \alpha) - \alpha, -\alpha \right) \\ &= \max \left(\frac{-C V_t + \left(\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t \right) (V_a + V_t - \alpha)}{C}, \alpha \right)\end{aligned}$$

Since $C > 0$, $\frac{-C V_t + \left(\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t \right) (V_a + V_t - \alpha)}{C} < 0$ is equivalent to:

$$\begin{aligned}&-C V_t + \left(\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t \right) (V_a + V_t - \alpha) < 0 \\ &-C V_t + \left(\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t (V_a + V_t - \alpha) \right) \\ &= -\frac{1}{2}\alpha V_t^2(1-k)^2 - \frac{1}{2}V_a^2 V_t(1-k)^2 - \frac{1}{2}V_a V_t^2(1-k)^2 - \alpha^2 k V_t(1-k) \\ &\quad - \frac{1}{2}(V_a V_t^2 - V_a \alpha^2 k^2 + \alpha^3 k^2 + \alpha V_t^2) - \Omega V_t(V_a + V_t)(1-k) - \Omega V_t \alpha k\end{aligned}$$

Thus, $\frac{-C V_t + \left(\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t \right) (V_a + V_t - \alpha)}{C} < 0$ is equivalent to $\Omega > \Omega_3$, where.

$$\Omega_3 = \frac{-\left(\frac{1}{2}\alpha V_t^2 + \frac{1}{2}V_a^2 V_t + \frac{1}{2}V_a V_t^2\right)(1-k)^2 - k\alpha^2 V_t(1-k) - \frac{1}{2}(V_a V_t^2 - V_a \alpha^2 k^2 + \alpha^3 k^2 + \alpha V_t^2)}{V_t((V_a + V_t)(1-k) + \alpha k)}$$

Because $\alpha < V_t$ and $k < 1$, Ω_3 is negative.

A.6. Proof of Corollary 2

Proof Both Ω_1 and Ω_0 are functions of α . For any given positive α , the merger is feasible if $\Omega \geq \Omega_1$ and $\Omega \geq \Omega_0$ from Proposition 4. That is, $\Omega \geq \max(\Omega_1, \Omega_0)$.

From Proposition 4, we know that if $\Omega < \Omega_0$, the merger is still feasible as long as $\Omega \geq \frac{-\frac{1}{2}V_a^2(1-k)^2 - V_a V_t(1-k)^2 - \alpha V_a k(1-k) + \phi V_t}{(1-k)(V_a + V_t) + \alpha k}$. This scenario implicitly specifies that $\Omega_1 \leq \Omega_0$. $\Omega_1 \leq \Omega_0$ is equivalent to:

$$\frac{(V_a + \phi)V_t}{V_a} < \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{V_t - \alpha}$$

The above inequality is equivalent to:

$$\begin{aligned} & \frac{1}{2}k^2V_a\alpha^2 + [(V_a + \phi)V_t + kV_aV_t(1-k)]\alpha \\ & + V_a\left(-kV_t^2 + \frac{1}{2}k^2V_t^2\right) - \phi V_t^2 \geq 0 \end{aligned}$$

Since $\alpha > 0$, the above inequality is equivalent to:

$$\alpha \geq \frac{-[(V_a + \phi)V_t + kV_aV_t(1-k)] + \sqrt{[(V_a + \phi)V_t + kV_aV_t(1-k)]^2 + 2k^2V_a(kV_aV_t^2 - \frac{1}{2}k^2V_aV_t^2 + \phi V_t^2)}}{k^2V_a}.$$

Appendix B: Deriving the Propositions Under the Case of $k \frac{V_a}{V_a + V_t - \alpha}$

If $k < \frac{V_a}{V_a + V_t - \alpha}$, $V_a + V_t - \alpha + \gamma$ will always be greater than $k(V_a + V_t - \alpha)$. In this case, the risk of early liquidation does not exist for the shareholders of the combined firm. The target shareholders' expected payoff, in a stock for stock exchange, from the acquisition, is then:

$$\frac{V_t - \alpha + \rho}{V_a + V_t - \alpha}(V_a + V_t + \Omega) \quad (19)$$

The target's shareholders will accept the offer if

$$\frac{V_t - \alpha + \rho}{V_a + V_t - \alpha} \cdot (V_a + V_t + \Omega) \geq \frac{1}{2}k^2 \frac{(V_t - \alpha)^2}{V_t} + (1-k)V_t + \alpha k \quad (20)$$

i.e., the premium ρ has to satisfy the following:

$$\rho \geq \alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{V_t(V_a + V_t + \Omega)} \cdot (V_a + V_t - \alpha) \quad (21)$$

The minimum premium demanded by the target decreases in Ω . That is, rational target would tradeoff merger premium and share of post merger expected gain in a stock for stock exchange; the higher the post merger gain is, the lower is the premium demanded by the target. We restrict the premium to be nonnegative, therefore the target demands the premium as, $\rho \geq$

$\max\left(\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{V_t(V_a + V_t + \Omega)}(V_a + V_t - \alpha), 0\right)$. That is, (1) if $\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{V_t(V_a + V_t + \Omega)}(V_a + V_t - \alpha) \geq 0$, the target demands the premium $\rho \geq \alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{V_t(V_a + V_t + \Omega)} \cdot (V_a + V_t - \alpha)$; (2) if $\alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{V_t(V_a + V_t + \Omega)}(V_a + V_t - \alpha) < 0$, the target demand the premium of

$$0. \quad \alpha - V_t + \frac{\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t}{V_t(V_a + V_t + \Omega)}(V_a + V_t - \alpha) \geq 0 \quad \text{is equivalent to } \Omega \leq \frac{(\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t)(V_a + V_t - \alpha)}{V_t(V_t - \alpha)} - V_a - V_t. \quad \text{Let us denote } \frac{(\frac{1}{2}k^2(V_t - \alpha)^2 + (1-k)V_t^2 + \alpha k V_t)(V_a + V_t - \alpha)}{V_t(V_t - \alpha)} - V_a - V_t \text{ as } \Omega_0.$$

The acquirer shareholder's expected value, in a stock for stock exchange, at date 2 is:

$$\frac{V_a - \rho}{V_a + V_t - \alpha}(V_a + V_t + \Omega) - \phi \quad (22)$$

ϕ is the acquirer's merger cost excluding the premium, for example, investment banking fees. ϕ is a positive number.

The acquirer shareholder's expected value, without the acquisition, at date 2 is V_a . Therefore, the acquirer will make the arbitrage motivated acquisition if

$$\frac{V_a - \rho}{V_a + V_t - \alpha}(V_a + V_t + \Omega) - \phi \geq V_a \quad (23)$$

That is, $\rho \leq V_a - \frac{(V_a + \phi)}{V_a + V_t + \Omega}(V_a + V_t - \alpha)$.

Because the premium is restricted to be nonnegative, the acquirer offers the following premium schedule:

If $V_a - \frac{(V_a + \phi)}{V_a + V_t + \Omega}(V_a + V_t - \alpha) > 0$, then $\rho \in \left[0, V_a - \frac{(V_a + \phi)V_t}{C}(V_a + V_t - \alpha)\right]$;

If $V_a - \frac{(V_a + \phi)}{V_a + V_t + \Omega}(V_a + V_t - \alpha) = 0$, then $\rho = 0$.

Proposition 6 *The necessary condition for the merger to be feasible is that $V_a - \frac{(V_a + \phi)}{V_a + V_t + \Omega}(V_a + V_t - \alpha) \geq 0$. That is,*

$$\Omega \geq \frac{V_a + \phi}{V_a}(V_a + V_t - \alpha) - V_a - V_t. \quad (24)$$

Proof If $V_a - \frac{(V_a + \phi)}{V_a + V_t + \Omega}(V_a + V_t - \alpha) < 0$, then the premium the acquirer is willing to offer has to be below zero, which violates the nonnegative premium constraint. Thus the merger is not feasible in that case.

Let us denote $\frac{V_a + \phi}{V_a}(V_a + V_t - \alpha) - V_a - V_t$ as Ω_1 . Ω_1 is a decreasing function of α . That is, the higher is the magnitude of undervaluation, the lower is the minimum value of post merger gain that is necessary for the acquirer to participate in the merger.

Proposition 7 *Given $\Omega \geq \Omega_1$, if $\Omega \geq \Omega_0$, the merger is feasible. Given $\Omega \geq \Omega_1$, if $\Omega \geq \Omega_0$, the merger is still feasible as long as $\Omega \geq \Omega_2 \equiv \frac{\frac{1}{2}k^2(V_t - \alpha)^2 - kV_t^2 + \alpha k V_t + \phi V_t}{V_t}$.*

The proof is similar to that of Proposition 4.

Corollary 3 For any magnitude of undervaluation, α , the merger is feasible as long as $\Omega \geq \max(\Omega_1, \Omega_0)$; when $\alpha \geq \alpha_1$, the merger is also feasible as long as $\Omega \geq \frac{\frac{1}{2}k^2(V_t - \alpha)^2 - kV_t^2 + \alpha kV_t + \phi V_t}{V_t}$, where

$$\alpha_1 = \frac{-(V_a + \phi)V_t + kV_aV_t(1 - k) + \sqrt{[(V_a + \phi)V_t + kV_aV_t(1 - k)]^2 + 2k^2V_a(kV_aV_t^2 - \frac{1}{2}k^2V_aV_t^2 + \phi V_t^2)}}{V_ak^2}.$$

The proof is similar to that of Corollary 2.

References

- Abreu, D., and M.K. Brunnermeier. 2002. Synchronization risk and delayed arbitrage. *Journal of Financial Economics* 66: 341–360.
- Almazan, A., K. Brown, M. Carlson, and D. Chapman. 2004. Why constrain your mutual fund manager? *Journal of Financial Economics* 73: 289–321.
- Barberis, N., and R. Thaler. 2002. A survey of behavioral finance. In *Handbook of the economics of finance*, ed. G. Constantinides, M. Harris, and R. Stulz. Amsterdam: North-Holland.
- Bris, A., W.N. Goetzmann, and N. Zhu. 2007. Efficiency and the bear: Short sales and the markets around the world. *Journal of Finance* 62: 1029–1079.
- Campbell, J., and A.S. Kyle. 1993. Smart money, noise trading and stock price behavior. *Review of Economic Studies* 60: 1–34.
- Cornelli, F., and D. Li. 2002. Risk arbitrage in takeovers. *Review of Financial Studies* 15: 837–868.
- Coyne, C., F. Fabozzi, and U. Yaari. 1992. How taxes transform corporate acquisitions into asset arbitrage. In *Research in finance*, ed. A. Chen, 173–203. Greenwich: JAI Press.
- D’Avolio, G. 2002. The market for borrowing stock. *Journal of Financial Economics* 66: 271–306.
- DeBondt, W., and H. Thompson. 1992. Is economic efficiency the driving force behind mergers? *Managerial and Decision Economics* 13: 31–44.
- Dechow, P.M., A.P. Hutton, and L. Meulbroek. 2001. Short sellers, fundamental analysis, and stock returns. *Journal of Financial Economics* 61: 77–106.
- DeLong, J.B., A. Shleifer, L.H. Summers, and R.J. Waldmann. 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98: 703–738.
- Dow, J., and G. Gorton. 1994. Arbitrage chains. *Journal of Finance* 49: 819–849.
- Duffie, D., N. Garleanu, and L.H. Pedersen. 2002. Securities, lending, shorting, and pricing. *Journal of Financial Economics* 66: 307–340.
- Fluck, Z., and A.W. Lynch. 1999. Why do firms merge and then divest? A theory of financial synergy. *Journal of Business* 72: 319–346.
- Gonzales, P., G. Vasconcellos, and R. Kish. 1998. Cross border mergers and acquisitions: The undervaluation hypothesis. *Quarterly Review of Economics and Finance* 38: 25–48.
- Gromb, D., and D. Vayanos. 2002. Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics* 66: 361–408.
- Jones, C.M., and O.A. Lamont. 2002. Short-sale constraints and stock returns. *Journal of Financial Economics* 66: 207–240.
- Lamont, O.A., 2003. Go down fighting: Short sellers vs. firms. Unpublished working paper, University of Chicago.
- Lamont, O.A., and R.H. Thaler. 2003. Can the market add and subtract? Mispricing in tech stock carve outs. *Journal of Political Economy* 111: 227–268.
- Mandelbaum, A. 1995. The source of gains in New Zealand takeover offers: Wealth creation or merely undervaluation. In *Research in international business and finance*, 12, ed. J. Doukas and L. Lang, 285–304. Greenwich: JAI Press.

- Merton, R. 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42: 483–510.
- Mitchell, M., and T. Pulvino. 2002. Characteristics of risk and return in risk arbitrage. *Journal of Finance* 56: 2135–2175.
- Mitchell, M., T. Pulvino, and E. Stafford. 2002. Limited arbitrage in equity markets. *Journal of Finance* 57: 551–584.
- Ofek, E., M.P. Richardson, and R. Whitelaw. 2004. Limited arbitrage and short sales restrictions: Evidence from the options markets. *Journal of Financial Economics* 74: 305–342.
- Shleifer, A., and R. Vishny. 1990. Equilibrium short horizons of investors and firms. *American Economic Review Papers and Proceedings* 80: 148–153.
- . 1992. Liquidation values and debt capacity: a market equilibrium approach. *Journal of Finance* 47: 1343–1366.
- . 1997. The limits of arbitrage. *Journal of Finance* 52: 35–55.
- Tuckman, B., and J. Vila. 1992. Arbitrage with holding costs: a utility-based approach. *Journal of Finance* 47: 1283–1302.
- Wurgler, J., and E. Zhuravskaya. 2002. Does arbitrage flatten demand curves for stocks? *Journal of Business* 75: 583–608.



Equity Premium Puzzle: The Distributional Approach

58

Nadezhda Safronova

Abstract

This chapter provides an extensive review of research on the equity premium puzzle, providing an overview of the existing approaches to this problem. We then propose an alternative approach including the application of the stable distribution. We analyze a portfolio optimization problem, under the assumption of normal (Gaussian) and stable (non-Gaussian) distributed asset returns. We examine and compare the results of portfolio allocations obtained in normal and stable cases. Finally, we investigate the ratio between the coefficients of risk aversion in normal and stable cases.

Keywords

Asset pricing · Equity premium puzzle · Habit formation · Preferences · Return distribution · Risk aversion · Stable distribution

58.1 Introduction

The term “equity premium” refers to the return on Stocks in excess of the return to a short-term “risk-free” instrument such as U.S. Treasury Bills. Mehra and Prescott (1985) first coined the term “Equity Premium Puzzle.” Their analysis, based on historical data for the 90-year period from 1889 to 1978, found that the equity premium exceeds estimates from standard asset pricing models by large magnitudes.

To provide insight as to the magnitude of the puzzle, note that real returns to U.S. government bonds average approximately 1% per year historically, while real returns to the U.S. equities average 7% per annum (Kocherlakota 1996). In

N. Safronova (✉)

School of Economics and Business Engineering, Institute of Econometrics, Statistics and Mathematical Finance, University of Karlsruhe, Karlsruhe, Germany

e-mail: nadezhda.safronova@statistik.uni-karlsruhe.de

comparison, standard economic models suggest the equity premium should be much smaller, in the range of 2–3% per year. Thus, general utility-based theories of asset prices have difficulty explaining why the magnitude of the equity premium observed from historical data is so great.

The “equity premium puzzle” brings to light an important issue in financial economics and challenges market efficiency. To illustrate, consider that if the return to equity investments is too high, why don’t investors allocate massive amounts of their wealth to stocks and reject low-returning bonds, thereby raising the price of stocks which will lower their subsequent returns? Thus, in equilibrium, the theoretical models do not accurately describe the actions of individuals collectively, suggesting that individuals base their decisions on different criteria.

To address the “puzzle,” the economics literature has focused on modifications to the formal utility-based theory and refinements and expansions of data sources. Despite much work in this area, the puzzle remains unsolved in that the plausible models still fail to wholly and adequately resolve the disconnect between the historical equity premium and estimates produced by the models.

Attempts to resolve the puzzle have become one of the major motive research in Finance and Economics. A number of generalizations to key features of the Mehra and Prescott (1985) model have been proposed in order to bring the models inline with the data. The following list summarizes these generalizations:

- **Preferences** (see Abel 1990; Bansal and Yaron 2000; Benartzi and Thaler 1995; Boldrin et al. 2001; Campbell and Cochrane 1999; Constantinides 1990; Epstein and Zin 1991; Ferson and Constantinides 1991),
- **Modified probability distributions considering rare but disastrous events** (see Rietz 1988; Mehra and Prescott 1988),
- **Data survival bias** (see Brown et al. 1995),
- **Incomplete markets** (see Bewley 1982; Brav et al. 2002; Constantinides and Duffie 1996; Heaton and Lucas 1997, 2000; Krebs 2000; Lucas 1994; Mankiw 1986; Mehra and Prescott 1985; Storesletten et al. 2001; Telmer 1993)
- **Market imperfections** (see Aiyagari and Gertler 1991; Alvarez and Jermann 2000; Bansal and Coleman 1996; Basak and Cuoco 1998; Constantinides et al. 2002; Danthine et al. 1992; Daniel and Marshall 1997; He and Modest 1995; Heaton and Lucas 1996; Luttmer 1996; McGrattan and Prescott 2001; Storesletten et al. 2001)
- **Attempts in modelling of consumers’ limited participation in the stock market** (see Attanasio et al. 2002; Brav et al. 2002; Mankiw and Zeldes 1991; Vissing-Jorgensen 2002),
- **Problems of temporal aggregation** (see Gabaix and Laibson 2001; Heaton 1995; Lynch 1996)
- **Behavioral explanations** (see Barberis et al. 2001; Mehra and Sah 2002).

Despite the numerous attempts at solving the equity premium puzzle, none of these attempts have fully resolved or explained the anomalies.

This chapter is organized as follows. Section 58.1 describes the initial model and assumptions set forth by Mehra and Prescott (1985). Sections 58.2 and 58.3 discuss numerous existing approaches to the equity premium puzzle, including the point of view that there is no equity premium puzzle at all. Section 58.4 describes the stable distribution, highlighting its advantages over the normal distribution in empirical data analysis. Sections 58.5 and 58.6 propose an alternative model, considering the stable distribution assumption.

58.1.1 The Lucas' Asset Pricing Model (CCAPM) and Standard Preferences

We begin with a description of the model set forth by Mehra and Prescott (1985). Their model is a variation of the pure exchange model of Lucas (1978). According to the model, there exists a maximization problem of an agent at date t , who has endowment each period, and decides whether to invest in some risky asset with price p_t and payoff x_{t+1} . The agent faces uncertainty concerning the date $t + 1$ payoff. Therefore, at time t , the agent faces their investment decision and bases that decision on their expectations, conditioned on all available information as of date t .

The formal description of the maximization problem is as follows:

$$\begin{aligned} \max_{\{\xi\}} & u(c_t) + E_t \beta u(c_{t+1}) \\ c_t &= e_t - p_t \xi \\ c_{t+1} &= e_{t+1} + x_{t+1} \xi, \end{aligned}$$

where e is the original consumption level, assuming the agent did not invest in the asset, and ξ is the amount of wealth the agent invests in the asset. The first order necessary condition is:

$$p_t u'(c_t) = E_t [\beta u'(c_{t+1}) x_{t+1}],$$

or equivalently may be expressed as any of the following asset pricing equations:

$$\begin{aligned} p_t &= E_t [m_{t+1} x_{t+1}] \\ 1 &= E_t [m_{t+1} R_{t+1}] \\ 1 &= E [mR], \end{aligned}$$

where the risky return, R_{t+1} , is $R_{t+1} = \frac{x_{t+1}}{p_t}$, and m_{t+1} is the *stochastic discount factor*. The stochastic discount factor is frequently referred to as the *pricing kernel* or *intertemporal marginal rate of substitution* since $m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}$.

The investor's preferences concerning random consumption are ordered by the increasing concave utility function U , which is restricted to be of the constant

relative risk aversion class in order to insure that the equilibrium return process is stationary. The individual's utility function is:

$$U(c) = \frac{c^{1-\gamma}}{1-\gamma},$$

where $c = c_{t+1}/c_t$ is the consumption growth, or one plus the growth rate, and γ is the coefficient of relative risk aversion, which influences greatly the risk premium. Mehra and Prescott (1985) assume that the growth rate of consumption and dividends are distributed log-normally.

The parameter γ measures the curvature of the utility function. When fitting the model to actual data, estimates of γ provide important information regarding how well the model represents the choice problem. Large estimated values of γ , those significantly exceeding 1, correspond to risk-seeking individuals and give rise to the question "why is the average equity return so high relative to the average risk-free rate." On the other hand, if γ is near zero, i.e. individuals are nearly risk-neutral, then the question becomes "why is the average equity return so high?"

Numerous studies calibrate the coefficient of risk aversion based on a series of questions designed to identify an individual's risk preferences. Based on these studies, it becomes evident that the coefficient of risk aversion should be a small number, certainly less than 10. Even in the marginal case, however, if the risk aversion coefficient is 10 and the discount factor is 0.99, the risk-free rate equals 12.7% and return on equity is 14.1%. This means that the equity risk premium is 1.4%, which is significantly lower than the historically observed equity premium (6.18%). Most studies find the optimal value of the risk aversion coefficient to be close to 3. Under those parameters, the risk-free rate will be even higher and the corresponding risk premium lower. Based on this evidence, consideration of various, plausible degrees of risk aversion fail to resolve the equity premium puzzle.

Researchers have found that the Lucas asset pricing model fits the empirical data well, especially in the context of macroeconomic research on Real Business-Cycle Theory. The model easily replicates the essential macroeconomic features of the business cycle (see Kydland and Prescott 1982). On the other hand, the model does not fit well the financial market data on stock returns, resulting in the equity premium puzzle.

Philippe Weil (1989) raises an additional puzzle, the **risk-free rate puzzle**. This puzzle pertains to the high risk-free rate generated by the high risk aversion coefficient value necessary for the model to produce a large equity premium. Under the assumption of log normally distributed growth rate of consumption and dividends, the risk-free rate may be decomposed into the following components:

$$\ln R_f = -\ln \beta + a\mu_x - \frac{1}{2}\alpha^2\sigma_x^2 + \varepsilon.$$

The first term, $-\ln \beta$, is a time preference or impatience term, i.e. $\beta < 1$ reflects the fact that agents prefer early consumption to later consumption. Therefore, the conditions of perfect certainty and no growth in consumption result in the unique interest rate in the economy: $R(f) = 1/\beta$.

The second term, $a\mu_x$, results from consumption growth. For high levels of expected future consumption, agents borrow against future consumption, aiming to smooth their lifetime consumption. Greater curvature of the utility function results in larger consumption growth rates, thus increasing the desire to smooth consumption. In the equilibrium this results in a higher interest rate since the aggregate agents cannot simultaneously increase their current consumption.

The third term, $\frac{1}{2}\alpha^2\sigma_x^2$, results from demand for risk-free saving. Facing uncertainty, agents seek to protect themselves against future unfavourable consumption levels by creating reserves of the consumption good. The result is that, in the equilibrium, the interest rate falls to as a result of the increasing demand for savings.

The fourth term, ε , is a risky term, describing the residual and the variance not explained by the model.

Hansen and Jagannathan (1991) provide an alternative perspective on the equity premium puzzle. They investigate the mean-variance frontier, which presents the maximum expected return for a given level of risk as measured by variance. All portfolios of risky assets lie inside the mean-variance frontier. The portfolios on the frontier present the optimal return-per-unit-risk and are correlated with each other and with the discount factor m .

The slope of the mean-standard deviation frontier $\sigma_t(m_{t+1})/E_t(m_{t+1})$ is the largest available ratio of mean excess return to standard deviation (Sharpe ratio):

$$\left| \frac{E_t(R_{t+1}) - R_t^f}{\sigma_t(R_{t+1})} \right| \leq \frac{\sigma_t(m_{t+1})}{E_t(m_{t+1})} \approx \gamma \sigma_t(\Delta c_{t+1})$$

where R_{t+1} is the risky return, $R_t^f = 1/E(m_{t+1})$ is risk-free return. This inequality reflects the equity premium puzzle and appears to be the Hansen-Jagannathan lower bound on the pricing kernel.

The slope of the mean-standard deviation frontier is larger if the economy is more risky, consumption is more volatile, or investors are more risk averse. The empirical post-war US data illustrate that the observed slope is much higher than suggested by reasonable risk aversion and consumption volatility estimates. With a Sharpe ratio of 0.5 and with $\gamma \sigma_t(\Delta c_{t+1}) = 0.01$, investors should have a risk aversion coefficient of $\gamma \approx 50$ or more, whereas various measures of risk aversion in the empirical data indicate a risk aversion below 10. According to this specification, the equity premium puzzle may result for the following reasons:

1. Investors are much more risk averse than expected;
2. Investors had good luck in the last 50 years, benefitting from a string of positive draws from the distribution of risky outcomes;
3. The model suffers from critical flaws.

Mehra and Prescott (1985), as mentioned above, use an extension of Lucas' (1978) asset-pricing model to estimate the premium for bearing systemic risk. Important to their analysis are the restrictions placed on the risk aversion parameter. Estimating the relative risk aversion parameter using statistical tools differs meaningfully from estimating the equity risk premium. Mehra and Prescott (1985) did not estimate the risk aversion parameter and did not estimate the discount rate parameters. Additionally, they reject extreme risk aversion based on observations on individual behaviour, which include the small size of premia for jobs with uncertain income and the limited amount of insurance against idiosyncratic income risk. Individuals with limited access to capital markets make investments in human capital resulting in very uneven consumption over time.

The basic growth model, when restricted to be consistent with the actual economic growth, generates business cycle fluctuations only if the risk aversion parameter is near zero. This corresponds to the log case in standard usage. Importantly, the risk aversion parameter appears in various observations at both the household and the aggregate levels and appears to be small. Mehra and Prescott find that for all values of the risk-aversion coefficient less than 10, which is decided to be the upper limit value for this parameter, a premium for bearing aggregate risk is not related significantly to the historic equity premium.

58.1.2 Equity Premium Puzzle in Different Countries

Historically, stock returns have been considerably higher than the relatively risk-free security. This result obtains from the estimates of the equity premium based on the various data sets used in the literature and application of consumption-based asset pricing models (C-CAPM). The U.S. equity market is not unique in terms of the large historical equity premium. Large equity excess returns are also observed in countries with sizeable capital markets, such as UK, Japan, Germany and France. Mehra and Prescott (2003) illustrate the equity premium in the post-war period for these countries using data based on Campbell (2001).

The annual return on the British stock market was 5.7% over the period of 1947–1999, whereas the average equity premium of 4.6% exceeds significantly the average return of 1.1% on a relatively risk-free security. They find similar results for Japan, Germany and France: average returns on risk-free securities and average equity premia reached 1.4% and 3.3%, 3.2% and 6.6%, 2.7% and 6.3% respectively for these countries. Moreover, empirical investigation of the Japanese markets during the period of 1980–1997 (Baba 2000) suggests that the risk-free rate puzzle is an even more compelling phenomenon than the equity premium puzzle in Japan during this period.

Engsted, Mammen, and Tanggaard (2001) investigate the Consumption CAPM (C-CAPM) and the equity premium puzzle using asset returns and consumption data from the US and Denmark. Their analysis finds large differences between US and Danish asset markets. Their results depend to some extent on the length of the investment horizon. It is worth noting that earlier work focuses attention on short

investment horizons (1 year). Engsted et al. (2000) investigate the effect of altering the horizon. The main results show that the evidence against the C-CAPM becomes much weaker when the horizon is extended from 1 to 5 years. The results for the Danish data show that risk-adjusted equity premia are insignificant even with low values of the risk aversion coefficient. However, according to tests of the model, it cannot be claimed that there is no equity premium puzzle evident. Indeed, the equity premium has varied considerably over time and the variation depends mostly on the time horizon over which researchers measure the premium. During certain periods, the equity risk premium varied to enough to produce negative premia.

58.2 Approaches to the Equity Premium Puzzle

Studies of the equity premium puzzle typically consider three specific assumptions. Each of these assumptions has been challenged in the literature, as we will describe in subsequent sections. These are:

1. *Beliefs and preferences.* Research assumes the specified utility function reflects the preferences of every investor. Although all investors may act as if they have the same utility function, this does not imply that the *stochastic discount factor*, m , is the same for everyone. Multiple values of the stochastic discount factor, m , makes the theory difficult to apply. Therefore, a tractable solution requires one, unique m . In one approach, it is assumed that all investors are identical, i.e. there is a representative investor. This restrictive condition may be relaxed by instead assuming that markets are complete.
2. *Complete markets.* Markets are complete when there exists, for each state, an asset with a payoff of one unit of wealth if that state occurs and zero if any other state occurs. The assumption that markets are complete does not mean that there is no uncertainty, but instead may be interpreted that agents are able to insure against every risky outcome. Under this assumption, it is possible to demonstrate that the stochastic discount factor is the same for all investors.
3. *Frictionless markets.* Most research makes the assumption that there are no transactions costs or constraints on asset trades. This unrealistic assumption may be a reason for the failure of the model to explain the observed equity premium.

58.2.1 Time and State Preferences

State dependent preferences have a long history in economics. Epstein and Zin (1991) present a class of preferences they refer to as “Generalized Expected Utility” (GEU) which allow for independent parameterization of the coefficient of risk aversion and the elasticity of inter-temporal substitution. In this framework, asset prices result from both the covariance with the growth rate of consumption and with the return on the wealth portfolio. This captures the pricing features of both the

standard consumption CAPM and the traditional static CAPM. Another feature of this class of models is that a high coefficient of risk aversion does not necessarily imply that agents want to smooth consumption over time. However, difficulty arises when testing this alternative preference structure involving variables that are unobservable, making calibration difficult. Epstein and Zin (1991) use the “market portfolio” to measure the wealth portfolio and suppose that their framework offers a solution to the equity premium puzzle. However it was found that their choice of proxy for the wealth portfolio exaggerates the correlation between asset returns and the wealth portfolio.

The preferences suggested by Epstein and Zin separate the coefficient of relative risk aversion (CRRA) from the elasticity of intertemporal substitution. Melino and Yang (2003) generalize the model of Epstein and Zin by allowing the representative agent to display state dependent preferences. Various combinations of state dependent CRRA with state dependent elasticity of intertemporal substitution were investigated. Under the condition of constant elasticity of inter-temporal substitution and time varying CRRA, the results are very similar to those generated without state preferences.

Another approach combines a state dependent elasticity of intertemporal substitution with a state dependent coefficient of relative risk aversion. In this approach, the model produces results that match the U.S. historical record. The additional restriction imposed by this class of preferences is that the coefficient of risk aversion depends strictly on the elasticity of inter-temporal substitution. Under this assumption, an agent averse to consumption variation across different states *at a particular point of time* is then averse to *consumption variation over time*. Mehra and Prescott (2003) find that there is no reason why this condition must be true. Since consumption is growing over time on average, the agents in the Mehra and Prescott (1985) study have no particular need to save. This modification can potentially resolve the risk-free rate puzzle under the log-normality assumptions and using the market portfolio as a representative for the wealth portfolio. It cannot, however, resolve the entire equity premium puzzle.

58.2.2 Habit Formation

Constantinides (1990) presents a new approach which is referred to as “habit formation.” According to habit formation, an agent’s utility is affected both by current and past consumption levels. Thus, habit formation captures a fundamental feature of human behaviour: that individuals adjust to a certain level of consumption and prefer strongly to avoid subsequent decreases in their consumption level. Two types of habit can be defined and are referred to as “internal” and “external” while there are two modelling perspectives, referred to as “difference” and “ratio.” Internal habit formation captures the notion that utility is a decreasing function of one’s own past consumption and marginal utility is an increasing function of one’s own past consumption. Models with external habit emphasize that the operative benchmark is the individual’s level of consumption relative to other agents in the economy.

Constantinides (1990) suggests a model with internal habit where utility is defined over the difference between current consumption and lagged past consumption.

According to the model, preferences make the agent extremely averse to consumption risk even when the level of risk aversion is small. Even if changes in consumption are small, changes in marginal utility may be large. Thus, while this approach can resolve the equity premium puzzle only with extreme aversion to consumption risk, it can be applied to the risk-free rate puzzle, since the resulting aversion to consumption risk increases the demand for bonds, whereas the risk-free rate is reduced. Furthermore, if the growth rate of consumption is assumed to be i.i.d., an assumption of this model is that the risk-free rate varies considerably over time. Since the model of Constantinides (1990) does not have an i.i.d growth rate, the model suffers from this problem.

An alternate approach to circumvent this problem has been proposed by Campbell and Cochrane (1999). Their model addresses the possibility of a decreasing state variable so that risk aversion varies in a highly nonlinear manner. Under reasonable parameter values, their model is able to generate a high equity premium, since the risk aversion of investors rises greatly when the chances of price declines become larger. Since risk aversion increases precisely when consumption is low, it generates a demand for bonds that results in lower risk-free rates. This model performs well when fit to both historical consumption and asset market data.

Abel (1990) presents another modification of the Constantinides (1990). Abel models a phenomena referred to as “keeping up with the Joneses.” The intention is to define utility of consumption relative to average per capita consumption. Thus, the model is an external habit model where preferences are defined over the ratio of consumption to lagged aggregate consumption.

The idea is that one’s utility depends not on the absolute level of consumption, but on one’s consumption level relative to others’. The implication is that an individual can become extremely sensitive and averse to consumption variation. Since equity may have a negative rate of return, the agent faces risk of decreasing personal consumption relative to others. Therefore, equity becomes an undesirable asset compared to bonds.

Abel (1990) defines utility as the ratio of consumption relative to average per capita consumption rather than the difference between the two. This is not a trivial modification (Campbell 2001) to preferences. While “difference” habit models can, in principle, generate a high equity premium, ratio models generate a premium that is similar to that obtained with standard preferences.

Otrok, Ravikumar, and Whiteman (1998 2001) present another model with habit preferences. They explore whether the introduction of habit preferences into the simple intertemporal consumption-based capital asset pricing model resolves the equity premium and risk-free rate puzzles. Otrok, Ravikumar and Whiteman demonstrate that agents with habit preferences care not only about overall volatility, but also about the temporal distribution of that volatility. Specifically according to the model, habit agents are much more averse to high-frequency fluctuations than to low-frequency fluctuations. The size of the equity premium in the habit model is intended to be determined by a relatively insignificant amount of high-frequency

volatility in U. S. consumption. Furthermore, the model's premium and returns are highly sensitive to changes in characteristics of the stochastic process for consumption. Unfortunately, the model also carries counterfactual implications for the equally dramatic changes in the equity premium and risk-free rate observed over the last 100 years.

As compared with the previous approaches, Otrok, Ravikumar and Whiteman employ spectral utility functions to examine an agent's attitude toward the temporal distribution of consumption volatility. When the agent's preferences show intertemporal non-separabilities (e.g. habits), the asset pricing model has important implications for the equity premium and the risk-free rate of return. To isolate the solving method, spectral utility is applied to the equity premium puzzle in the habit model. The habit model produces the equity premium by showing extraordinary sensitivity to high-frequency fluctuations. Due to this sensitivity, as the volatility and serial correlation properties of U.S. consumption have changed during the last 100 years, the model makes counterfactual predictions of the time path of the equity premium and the risk-free rate. Therefore, important questions persist as to whether habit formation models can resolve the equity premium puzzle.

58.2.3 Idiosyncratic and Uninsurable Risk

Constantinides and Duffie (1996) develop a model with heterogeneous consumers to reflect the notion that consumers are dependent on idiosyncratic risk against which they are unable to insure themselves fully. Their model considers the consumers unable to insure against the risk of job loss or other major personal disasters. Equities and related investments possess the undesirable feature that they drop in price when the probability of job loss increases. In these situations, consumers require an extra incentive to hold equities and similar investment instruments. The equity premium results as the necessary enticement for these investors to hold risky equities. The model generates a high risk premium, but may not generate the required degree of consumption variation generated in an economy with agents who have a relatively low level of risk aversion.

Lettau (2001) evaluates models with idiosyncratic consumption risk using Hansen and Jagannathan's (1991) volatility bounds. Idiosyncratic risk helps to enter the bounds when this risk depends on the aggregate state of the economy. The model, however, does not pass the Hansen and Jagannathan test even for very volatile idiosyncratic income data. Therefore, Lettau's results suggest that the models with idiosyncratic uncertainty do not solve completely the equity premium puzzle.

58.2.4 Disaster State and Survival Bias

Rietz (1988) proposed a new approach to the equity premium puzzle. This approach considers scenarios having a very small probability of a very large drop in

consumption. Rietz finds that in such a scenario the risk-free rate is much lower than the return on equity. To produce this result, the model requires a 1% probability of a 25% decrease in consumption and a risk aversion parameter of 10. Such a scenario has not been observed in the USA for the recent years during which economic data exist. However, such a scenario is unlikely to be observed so frequently that this approach could be put into practice.

Another attempt at resolving the puzzle comes from Brown et al. (1995) and focuses on survival bias. It is particularly important that measured returns reflect accurately the true premium earned by stock market investors. Many stock exchanges were unsuccessful, resulting in low expected equity premia. Since it was not known in advance which exchanges would survive, stock and bond markets must be differentially influenced by a financial crisis. Mehra and Prescott (2003) find that although a survival bias may influence the levels of both the return on equity and debt, there is no evidence to support the assumption concerning the differential influence on the returns to stocks and bonds. Therefore, the disaster state and survival model is not a solution to the equity premium puzzle.

58.2.5 Borrowing Constraints

Constantinides, Donaldson, and Mehra (1998) present a new perspective on the equity premium puzzle with an overlapping-generations (OLG) economy in which a representative agent's life is divided into three overlapping periods. During the first period, the agent receives a relatively low endowment income. During the second period the agent is employed and is exposed to large wage income uncertainty. Finally, during the third period the agent retires and consumes the assets accumulated in the second period. This approach is an attempt to address such questions as: (1) why is the historical equity premium so high; (2) why the real rate of interest is so low; and (3) why so few consumers hold equities and bonds.

The researchers consider two versions of the overlapping generations economy:

1. Borrowing-constrained version.
2. Borrowing-unconstrained version.

In the case where the agent faces borrowing-constraints, the equity premium exhibits low correlation with young workers' expected wage income during their middle ages. This is in contrast to the borrowing-unconstrained economy where the young agent borrows and invests in equities, leading to reduction of the mean equity premium while the interest rate rises. Therefore, the borrowing-unconstrained models suffer from difficulty explaining the high mean equity premium and the low rate of interest.

In the overlapping generations model, a natural question is "who invests in bonds if equity is so attractive to the young wage earners?" Since the savings oriented middle-aged experience uncertainty in their future wage income and the majority of their future income derives from their current savings in equity and bonds, they

therefore prefer to hold a diversified portfolio that includes positive holdings of bonds, thus explaining the demand for bonds. The main features captured by the overlapping generation economy are that the future income of the young agents obtains from their wages during middle age, while the future income of the middle-aged agents derives from their savings through holdings of equities and bonds. As a result of this difference, the young have a distinct attitude to the mean equity premium compared the middle-aged.

As compared with equilibrium models, the OLG models have some important advantages. Equilibrium models that allow for consumer bond holdings, consistent with infinitely-lived consumers considered in some models, are difficult to construct. By contrast, in the OLG model the borrowing constraint plays a very different role in the consumers' life-cycle demand for equity and bonds.

The main contribution of the Constantinides, Donaldson, and Mehra (1998) paper is that the borrowing-constraint results in a lower interest rate and larger equity premium. On the other hand, in the case with incomplete markets, none of the traded securities correspond to bonds and the equity premium cannot be computed. Furthermore, the model does not resolve the equity premium puzzle since increasing the equity and bond return results in a lower equity premium. Absent borrowing constraints, the demand for bonds reduces and the risk-free rate puzzle again presents.

58.2.6 Transaction Costs

An important assumption in the traditional model is that the agent may rebalance their portfolio, without cost, as many times as desired over their investment horizon. The presence of transaction costs implies that investors may reduce the frequency at which they transact, for example they may choose to rebalance only once every several years.

The combination of infrequent portfolio trades and high returns on equities over long periods makes it plausible that, for longer horizon, the larger are investors' portfolio allocations to equities. However, although equities offer high expected returns over long periods, they also bear the possibility of large losses. Thus, the agent's decision depends on their preferences towards risk. More risk averse investors likely hold smaller shares of equities regardless of the frequency at which they rebalance their portfolio holdings. The transaction costs approach does not resolve entirely the equity premium puzzle. At best, the transactions cost approach explains a few, select aspects of the puzzle.

58.2.7 Taxes

McGrattan and Prescott (2001) propose an explanation for the equity premium based on changes in tax rates. It is important that tax rate explanation is not risk-based so that an equity premium can be taken into account but not as an equity *risk* premium.

McGrattan and Prescott find that, at least in the post-WorldWarII period, the equity premium is not puzzling after taking into account tax rates. They find that the significant decline in individual income tax rates and the increased opportunity to shelter income from taxation resulted in a doubling of equity prices between 1960 and 2000 and, therefore, the high observed returns on equity as compared to returns on debt. This argument can be illustrated by use of a simple one-sector model that includes only taxes on corporate distributions and taxes on corporate profits. Every agent seeks to maximize utility subject to their budget constraint, which represents the condition that the present discounted value of charges must be less than or equal to the discounted value of after-tax income. Agents' bear costs for their consumption and transactions of stocks. Agents receive income from three sources – dividends, wages, and government transfers. Agents pay personal taxes on dividend and wage income. The result of the study by McGrattan and Prescott (2001) is that a large increase in the personal tax rate with little change in the corporate tax rate raises indeed the value of equity. However, to show this interdependence, some specific conditions should be observed which are transitive.

Mehra and Prescott (2003) find another potential problem with investigating the McGrattan and Prescott model, which arises because of the difficulty in identifying the marginal investor. The marginal tax rate and the rate that applies to the marginal investor may differ greatly, while the rate that applies to the marginal investor is relevant for asset pricing. McGrattan and Prescott obtained the estimate of this tax rate from the weighted average rate calculation, the reliability of which is, however, questionable.

58.2.8 Incomplete Markets

Researchers have long recognized that incomplete market structures are a potential explanation of the equity premium puzzle. The analysis of incomplete market structures has been conducted using different empirical methodologies, such as simulation techniques. Whereas some early investigations note that incomplete markets cannot solve asset pricing puzzles, it is now widely accepted that the presence of uninsurable risk in a model economy results in model estimates that more closely resemble the predicted asset returns observed in practice. The results from models where markets are not assumed to be complete are sensitive to a number of parameters that have to be measured before setting up the model to explain asset returns.

Jacobs (1999) presents one interpretation of the equity premium puzzle and the risk-free rate puzzle by focusing on the importance of incomplete market structures and on the measurement of the pricing kernel. Early papers on incomplete markets place large emphasis on asset pricing puzzles and less on statistical tests of the model. These tests are of great interest for asset pricing as their interpretation is sometimes problematic due to data limitations and the testing technique.

Jacobs (1999) investigates a different approach to the measurement error problem. It is important to consider the measurement error because it helps to lower the

level of risk aversion. Jacobs' approach to analysing market incompleteness focuses on the restrictions implied by the model under a weaker set of assumptions, such as the following inter-temporal Euler equations

$$1 = \beta E_t \left(\frac{c_{i,t+1}}{c_{i,t}} \right)^{\gamma-1} R_{r,t,t+1}$$

$$1 = \beta E_t \left(\frac{c_{i,t+1}}{c_{i-t}} \right)^{\gamma-1} R_{ri,t,t+1}$$

where $c_{i,t}$ is the consumption of agent i in period t , $R_{r,t,t+1}$ and $R_{ri,t,t+1}$ are the risk-free and the risky return between periods t and $t + 1$, β is the discount factor, E_t is the mathematical expectation conditional on information available at t , and $1 - \gamma$ is the rate of relative risk aversion.

In the case where markets are incomplete, there are a maximum of T observations on N households. The solution to the optimization problem for these household data results from a generalized method of moments (GMM) framework. To construct the statistic for the test, assume that the Euler equation errors for a household are uncorrelated over time. The correlation between different households at the same time is determined by the size of the "aggregate shock" influencing the economy at that time, and therefore may be nonzero.

The paper demonstrates that the incomplete markets setting may plausibly lead to resolution of a number of important issues in asset pricing, such as the equity premium puzzle and the risk-free rate puzzle. One of the potential problems with tests of incomplete markets, however, is the presence of substantial measurement error. While the rate of risk aversion implied by the Euler equations is affected, indication that measurement error is important, the level of risk aversion is still intuitively plausible for asset-holders. These conclusions support a number of notes in the existing literature on incomplete markets that are typically obtained using simulation techniques. The conclusion of this analysis is that inclusion of uninsurable risk can adjust better the predicted asset returns to those found in the historical record, but often results in new problems tied to other parameters.

58.3 Is There the Equity Premium Puzzle?

58.3.1 Adjusted Set of Measurements

McGrattan and Prescott (2003) argue in favour of revisions to the approach set forth by Mehra and Prescott (1985). In particular, they make the following suggestions:

1. The T-Bill rate should not be used as the 'risk-free' rate since most households hold long-term debt in their portfolios instead of short-term government paper;
2. The costs of holding diversified equity portfolios should be taken into account;

3. Taxes on dividends should be calculated on the base of the equity portfolio returns;
4. It should be considered that the equilibrium asset pricing condition, used by Mehra and Prescott (1985), did not hold during WWII and the Korean War as the government imposed restrictions on production, consumer credit, and financial market fluctuations.

Using long-term high-grade bonds (and municipal bonds) as the ‘risk-free’ instrument, and making adjustments (2) and (3) to equity returns (and abstracting from the regulated sub-period 1935–1960) results in average excess real returns less than 1%. In other words, the ‘average equity premium puzzle’ is no longer a puzzle under these modifications.

Imrohoroğlu (2003) use the measurements proposed by McGrattan and Prescott (2003) and examine various aspects of the equity premium puzzle in detail, including the related low risk-free rate and excess volatility puzzles. The result of the research are support for two of the statements by McGrattan and Prescott (2003). Their adjustments do in fact solve the “average equity premium puzzle” and come close to solving the “risk-free rate puzzle.” However, the low risk-free rate puzzle, and the excess volatility puzzle persist. Standard theory cannot take account for the excess volatility of equity returns relative to consumption growth and the low correlation of excess returns with aggregate consumption growth. An alternative point of view held by a group of academicians and professionals is that currently there is no equity premium and, by implication, no equity premium puzzle.

58.3.2 Statistical Data Analysis

One of the main discussion points in the equity premium puzzle literature is the reliability of the data sources for the early studies conducted on this topic. Mehra and Prescott (1985) subdivide the U.S. historical equity premium data into three sub-periods. During the first subperiod, spanning the years 1802–1871, the equity return data are not particularly reliable for multiple reasons. First, the data are constructed of different market indexes and portfolios. Second, different numbers of stocks were used for the construction of the indexes during the different periods. Further, the structure of the indexes is not equivalent. Third, important details were not considered during index construction, such as the average of the bid and ask prices were used, ignoring transaction costs, and dividends were omitted from returns computations (Schwert 1990; Macaulay 1938).

During the second subperiod, which covers the years 1871–1926, the equity return data are also constructed from multiple sources, including newspapers and industrial portfolios. Unlike the indexes for the earlier sample period which ending in 1871, the return computations account for dividend distributions, so that these data better reflect true returns, although they are not themselves completely reliable. During this time period, there was no definitive source for the risk-free

rate. Thus, Mehra and Prescott (1985) make critical assumptions regarding the index construction.

The equity return data during the third subperiod, beginning in 1926 are of the highest quality, owing to the fact that the NYSE and the Centre for Research in Security Prices databases provide more accurate financial data. Clearly there exist differences in data quality across each period, which suggests caution when interpreting results comparing results between the different time periods.

58.3.3 Merton's Model: Conflict with the Equity Premium Puzzle

Bossaerts (1995) emphasizes several facts that contradict the results of the seminal paper by Mehra and Prescott (1985). According to the results of Mehra and Prescott, they substitute the smooth per capita consumption on non-durables and services for the more variable true payment process when calculating the price of the market index. In fact, their theoretical equity premium constitutes a lower boundary for the actual premium. Hence, unlike the Mehra and Prescott (1985) conclusions, they do not reject Lucas's equilibrium model.

Mehra and Prescott (1985) compare the average risk premium (the difference between the average return on the market and the risk free rate) observed during the period from 1889 to 1978 against the calculated values using a general equilibrium model. Merton (1980), interested in estimating the expected return on the market, used an equilibrium model similar to the one used in Mehra and Prescott (1985). Both models are based on a Markov-type economy with a representative investor (agent) having a constant relative risk aversion utility function. Merton's model relates the equity premium to the variance of the market return:

$$R^e - R^f = \alpha \sigma^2$$

where R^e is the expected return on the market, R^f is the risk free rate, σ^2 is the variance of the market return and α is the investor's relative risk aversion. From Merton's investigation of monthly returns of the New York Stock exchange Index over the period of July, 1926–June, 1978, one can calculate a relative risk aversion equal to 1.89 (see Merton 1980). The equity premium Merton observed over this 52 year period is close to the one in Mehra and Prescott's article. However, Mehra and Prescott's model implied a reasonable premium value, if one allows the relative risk aversion to be as high as 10. Mehra and Prescott's results clearly reject the equity premium model, whereas Merton's findings, in contrast, support the model.

Bossaerts (1995) finds that there was another mismatching of Mehra and Prescott's calculations. They state, that the sensitivity to δ , the standard deviation of the consumption growth rate, is larger. The average equity premium was roughly proportional to δ squared. To the contrary, Bossaerts argues that this runs contrary to the Modigliani-Miller dividend irrelevance theorem which states that the price of a security does not depend on the payout policy chosen by the firm. Mehra and

Prescott generate this variance sensitivity by assuming that consumption equals output and dividends for all times t . Merton's model implies the same variance sensitivity: the equity premium is proportional to the variance of the market return. Therefore, Mehra and Prescott find the average equity premium over the period 1889–1978 to be much higher than the one implied by their model and the per capita consumption on non-durables and services over that period. Their test investigates the lower boundary on the equity premium. Using Mehra and Prescott's model, but restricting δ to equal the standard deviation of the market return, and remembering that the equity premium is roughly proportional to δ squared, it follows that the equity premium should at most be well above the observed level. This is the corresponding upper boundary test. Hence, one can conclude that the average risk premium for the period 1887–1978 lies somewhere between a lower boundary and an upper boundary. It does not contradict either boundary implied by the Arrow Debreu general equilibrium model. To get a specific value for the required equity premium, Merton's method can be used. And his results for the period 1929–1978 indicate that the average equity premium is not inconsistent with the equilibrium model.

58.3.4 Rational Beliefs

Kurz and Beltratti (1996) determine that the equity premium puzzle is no puzzle at all by suggesting an approach utilizing the theory of rational beliefs equilibrium (RBE). In an RBE, agents require compensation for the uncertainty which is not permitted under rational expectations as in Mehra and Prescott (1985). The main idea of the approach is replacing of the theory of rational expectations with the theory of rational beliefs of which rational expectations as a special case. A number of tests of the rational expectations theory were presented by Kurz and Beltratti with a conclusion that this theory failed the tests permanently. As a conclusion, the theory of rational beliefs was applied to provide an intuitively simple explanation for the data which led to the equity premium puzzle.

The theory of RBE proposes that agents do not have “structural knowledge” about the economy, i.e. they do not know the correspondence between exogenous variables and asset prices or the true probability distribution of the dividend process. The component of prices attributing to the beliefs of the agents is *endogenous uncertainty*. It is the price uncertainty which the agents face in the market.

The equity premium calculations of Mehra and Prescott (1985) assume that only capital gains and losses which can be attributed to the variability of the exogenous variables factor into the risk faced by the agents. The implication of this assumption is that the Mehra and Prescott equity premium is calculated under the assumption of rational expectations of the agents, therefore excluding all the endogenous uncertainty. Kurz and Beltratti (1996) insist that the endogenous uncertainty is the dominant form of uncertainty in the equities markets.

58.3.5 The Equity Premium Puzzle, Looking Forward

In response to recent research's challenges to the existence of the equity premium puzzle, Mehra and Prescott (2003) stress the distinction between different interpretations of the term "equity premium." The first is the *ex post*, or realized, equity premium which is the actual, historically observed difference between the return on the market as reflected by a stock index, and the risk-free rate, provided by the return on government bills. This definition of the equity premium corresponds to the premium puzzle addressed by Mehra and Prescott (1985). The second, related concept is the *ex ante* equity premium, a measure of the predicted premium, i.e. the equity premium that is *expected* to prevail in the future, or the conditional equity premium.

The choice of one of these interpretations of the equity premium is important for an investment advisor and depends on the planning horizon. The equity premium documented by Mehra and Prescott (1985) reflects very long investment horizons. The *ex post* equity premium is not useful as a forecast of the premium for the next few years, since it is the realization of a stochastic process over a defined period of time, and has varied considerably over time.

Furthermore, the variation in the realized premium depends on the time horizon over which it is measured.

Investors planning over short-term investment horizons wish to project the conditional, expected equity premium over that planning horizon. Even if the current equity premium is small, this does not imply that the historical premium was too high or that the equity premium has declined. Mehra and Prescott emphasize that researchers should not dismiss the equity premium puzzle, and emphasize the need to first understand the observed phenomena, and then obtain a plausible explanation why the future is likely to differ from the past and why one may not expect the equity premium puzzle to persist in the future. Absent such an explanation, Mehra and Prescott (2003) suggest that, on the basis of our current collective knowledge, that over the long term the equity premium is likely to be repeated in the same way that it has been in the past. The failure of financial and economic modelling to capture the characteristic that appears to make stocks relatively risky puts the vital capacity of using this class of models in question.

58.4 Stable Distributions

Investors base their financial investment decisions almost exclusively on the expected returns and risk associated with the initial investment possibilities. The distributive character of the returns is of great importance for the theoretical and empirical analysis in economics and finance. For instance, the portfolio and the option pricing theory are based on assumptions regarding distribution functions. Previous researchers have advocated the application of the stable distributions in portfolio optimization. Early theories proposed by Bachelier (see Clark 1973) considered normal distribution assumptions. Other researchers, however, doubt in

the suitability of the normal distribution (see Mandelbrot 1963a, b; Mandelbrot and Taylor 1967; Fama 1965). In their research they find additional kurtosis, i.e. high peaks of the density function, which reject the normal distribution assumption. As an alternative, they suggested the stable distribution. Further research has confirmed this assumption (see, among others, Teichmoeller 1971; Officer 1972; Mittnik et al. 1997). Balke and Fomby (1994) demonstrate that most of the economic time series exhibit non-Gaussian behaviour. Their analysis of 50 U.S. macro-economical series, includes the Consumer Price Index (the measure of price inflation), industrial nominal compensation per hour (measure of wage growth), yields on AAA bonds, S&P 500 stock price index and GNP deflated. According to their investigation, the residuals of the autoregressive models point at the obvious presence of excess kurtosis and skewness. They conclude that even after extending the analysis to include GARCH class models, excess kurtosis and/or skewness remain.

Ziembra (1974) uses the symmetric stable distribution in an approximating problem of portfolio optimisation. Unfortunately theoretical and computational difficulties have decreased the interest in this approach. However, theoretical research and the development of computer algorithms in recent years have stimulated the intensive use of the stable distributions. Recent research by Ortobelli, Huber, Rachev, and Schwartz (2002), Ortobelli, Huber, and Schwartz (2003), and Ortobelli, Rachev, Huber, and Biglova (2003) compare the optimal distributions of risky and risk-free assets under alternative hypotheses of normal and stable distributed asset returns. In both one- and multi-period analysis, the results highlight that the assumed distributions result in up to 85% differences.

Monographies by Samorodnitsky and Taqqu (1994) and Janicki and Weron (1994) provide detailed descriptions of stable distributions properties along with numerical methods and computer simulations of generated random stable distributed variables and processes. We summarize briefly the characteristic of the stable distributions.

Compared to the normal distribution, stable distributions possess such features as heavy tails and high peaks of the density function relatively to the center and asymmetry is possible.

A stable distribution is characterized by the following parameters:

- α – index of stability or, characteristic parameter;
- β – skewness parameter of the density function;
- σ – scale parameter;
- μ – location parameter.

The parameter α controls the behaviour of the distribution tails. Stable distributions act the same way under infinity as the Paretian distributions. In these terms, the tails of stable distributions correspond to the Paretian type. The skewness (asymmetry) parameter, $\beta \in [-1, 1]$, reflects the extent of distributional asymmetry. If $\beta = 0$, then the distribution's density function is symmetric. If $\beta > 0$, the density function is skewed to the left, and the extent of skewness increases while β value comes closer to 1. The case when $\beta < 0$ corresponds to skewness to the right. Parameter

σ is scale parameter. In case of normal distribution ($\alpha = 2$) $DX = 2\sigma^2$. Note here that dispersion is $2\sigma^2$, instead of σ^2 , as compared with standard considerations. If $\alpha < 2$, dispersion DX does not exist. Parameter μ is location parameter, since in case of $\alpha > 1$ mathematical expectation $E|X| < \infty$ and $\mu = EX$. Generally, this interpretation is lacking, since EX can be not determined.

Summarizing the section, it is worth noting that stable distributions possess two attractive features:

1. Stable distributions have a domain of attraction; thus in this area the distributions of the observed values are also stable, or, rather they are close to the stable distribution features. Therefore, the “idealized” stable distribution does not distort the empirical distribution.
2. Each stable distribution has an index of stability, which does not depend on the observed time horizon. Index of stability can be considered as a general parameter, which can be applied for inference.

Stable distributions and stable processes are of significant importance as natural and likely candidates within construction of probability models of distributions and the evolution of financial indexes (e.g. currency rates, asset prices etc.). Although the application of the stable distributions in portfolio optimisation is not a new idea, theoretical and computational difficulties coupled with the theoretical failure in adjusting the empirical data to the models have hindered the advancement of this approach. Hence, theoretical research and computer technique development in recent years have stimulated the intensive use of the stable distributions.

58.5 Risk Measure

Value-at-Risk (VaR) is a widely used risk measure in applied finance (see Duffie and Pan 1997), answering the question “what is the maximum loss corresponding to a specified confidence level?” There exist various methodologies for the modelling of VaR. For an ordered confidence level α of $(0; 1)$ and time horizon Δt the measure of risk VaR estimates the maximum loss in the time period Δt with the confidence level α . If $f(R^p)$ is the distribution of the return losses for the final portfolio return, then 99% VaR is defined as 1% of the lower quantile of the return distribution. If the investor searches a compromise between average final return and 99% VaR of the final return, the criterion function can be defined as:

$$U(R^p) = E(R^p) - cVaR_{99\%}(R^p). \quad (1)$$

See Huisman et al. (1999) and Basak and Shapiro (2001) for further details.

Although VaR is a very popular measure of risk, it has undesirable properties such as lack of sub-additivity (i.e. the risk of a portfolio with two instruments can be greater than the sum of individual risks of these instruments), non-convexity and non-smoothness Artzner et al. (1999) suggest an alternative measure of risk with

more attractive properties named Conditional Value-at-Risk (CVaR), which is defined in the following way:

$$CVaR_{99\%} = E[E^p | R^p < VaR_{99\%}(R^p)]. \quad (2)$$

CVaR may be optimised using linear programming and non-smooth optimisation algorithms, which allows for implementation across portfolios with very large numbers of instruments and scenarios. Numerical experiments indicate that the minimization of CVaR also leads to VaR optimisation, since CVaR is always greater than or equal to VaR.

58.6 The Model

The analysis considers a portfolio consisting of two types of assets: a risky asset and a relatively risk-free asset, such as Treasury Bill.

The investor aims to optimise their portfolio allocation to achieve the maximum utility. At the same time, it is necessary to invest all the money he has in the available assets. The share (weight) of each asset in the portfolio should be non-zero.

In the present work this problem is solved in a discrete statement, i.e. when there is a final set of admissible portfolios among which the investor makes their choice. The solution in this case results from an exhaustive search of all feasible portfolios in each definite case.

Apart from the problem of utility function maximization, there is the problem of the optimal risk aversion coefficient which describes the investor's attitude to risk. A large value of the coefficient means the investor exhibits high aversion to risk, requiring large expected returns to entice them to take additional units of risk.

The analysis assumes normal and stable asset returns distributions. The aim of this research is to define for which type of asset returns distribution (normal or stable), the decision on asset allocation approximates more closely the empirical data.

The aim of the investigation is to demonstrate the generalized results of the distributional assumptions.

Let us indicate the following parameters:

- p – number of the portfolio, $p = 1, \dots, 100$;
- $z_{i,t}$ – price of the asset i at moment t ;
- $r_{i,t}$ – return of the asset i at moment t ;
- $x_{i,p}$ – simulated weight of the asset i in portfolio p ;
- R_p – return of portfolio p ;
- $E(R_p)$ – mathematical expectation of portfolio return;
- $CVaR_{\varepsilon}(R_p)$ – measure of portfolio risk under the error level ε ;
- ε – error level, $(1 - \varepsilon)$ – confidence level;
- c – coefficient of relative risk aversion;
- $U(R_p)$ – portfolio utility function.

Formally, the problem can be represented in the following way:

$$\max U(Rp_t) = E[Rp_t] - c \cdot CVaR_e(Rp_t), \quad (3)$$

where the mathematical expectation of the portfolio return at moment t is calculated as follows:

$$E[Rp_t] = \frac{1}{N} \sum_{n=1}^N Rp_{n,t}. \quad (4)$$

Portfolio return is defined as

$$Rp_t = \sum_{i=1}^d x_{i,t} \cdot r_{i,t}, \quad (5)$$

where d is the number of the assets in the portfolio.

Portfolio risk or, the portfolio loss level is defined by the measure of risk $CVaR_e(Rp)$ [52].

The admissible values of the coefficient of relative risk aversion are $c > 0$.

Since the values of $E[Rp_t]$ and $CVaR_e(Rp_t)$ do not depend on the coefficient of relative risk aversion, the utility function of every portfolio as a function of risk aversion coefficient is linear:

$$U(c) = k \cdot c + b, \quad (6)$$

where $U(c) = U(Rp_t)$; $b = E(Rp_t)$; $k = -CVaR_e(Rp_t)$, i.e. the plot of the utility function as a function of the risk aversion coefficient is represented as a straight line. Therefore, the plot of the optimal (maximum) utility function is the upper boundary of a set of these lines, i.e. is formed by the sections of these lines, where the corresponding line is the highest.

58.7 Calculations and Results Analysis

The analysis considers the following portfolios:

1. DowJonesIndustrials index and 3-monthT-bill for the period of January 02, 1997–June 30, 2003 (1563 days);
2. 24 assets included to the DAX30 index for the period of January 02, 1998–June 30, 2003 (1387 days);
3. DAX30 index and 6-month EURIBOR for the period December 28, 1998–June 30, 2003 (1139 days);
4. DAX30 and CAC40 indexes and 6-month EURIBOR for the period December 28, 1998–June 30, 2003 (1146 days).

A historical vector of portfolio weights for October 10, 2004 was accepted as an empirical portfolio in the second case (the source of the data was Deutsche Börse AG).

For the cases 1 and 3, the standard vector of weights for portfolios including one risky and one risk-free asset was accepted as an empirical portfolio: 0.4 for risk free assets and 0.6 for risky assets. In case 4 the empirical optimal portfolio was defined in the following way: 0.4 is the weight of risk-free asset (Euribor), the rest 0.6 points for the risky assets were divided as 0.36 for CAC40 and 0.24 for DAX30, since it is empirically decided to consider the French assets 1.5 times more reliable than the German assets.

The plots of the density functions' histograms for simulated and empirical asset returns demonstrate that:

- The density functions of the empirical asset returns exhibit heavy tails, while the simulated normal asset returns do not. At the same time, the simulated stable asset returns are closer to the empirical data in the sense that the density function in the stable case also has heavy tails.
- The peak of the density function of the empirical asset returns is higher than in the normal case, while the peak of the density function in stable case also stretches across the area of central values.

The utility function's dependence on risk aversion coefficient was investigated by means of constructing the following linear regression function, which characterizes the correspondence between the difference in values of the utility functions in normal and stable cases, $U_{normal}(c)$ and $U_{stable}(c)$, respectively, and the risk aversion coefficient c :

$$U_{stable}(c) - U_{normal}(c) = b_0 + b_1 \cdot c + \xi_1. \quad (7)$$

The numerical results for the portfolio of DowJonesIndustrials index and 3-monthT-bill show that the predicted value of the utility functions difference have the following view:

$$U_{stable}(c) - U_{normal}(c) = 0.0087 + 0.8206 \cdot c, \quad (8)$$

where the p -level (or, statistical significance) is close to zero (0.037), which indicates that there is a small probability of error of assuming that the calculation result is "true."

The results given in the Table 1 results from extending the analysis described above to the additional time series. The results presented below describe the utility function's dependence on the risk aversion coefficient for the mentioned above portfolios.

The final analysis investigates the ratio of the risk aversion coefficients in normal and stable cases, where the optimal risk aversion coefficient (under condition of maximum utility and closeness of the simulated weights to the empirical weights) in

Table 1 Utility function dependence on risk aversion coefficient

Portfolio	Period, days	$U_{stable}(c) - U_{normal}(c) = b_0 + b_1 \cdot c + \xi_1$		
		b_0	B_1	$p\text{-level}$
DowJonesIndustrials and 3-monthT-bill	1563	0.0087	0.8206	0.037
24 assets of DAX30	1387	0.0563	0.6972	0.093
DAX30 and 6-month EURIBOR	1139	0.0098	0.9273	0.016
DAX30, CAC40 and 6-month EURIBOR	1146	0.0352	0.9447	0.0047

Table 2 Dependence between risk aversion coefficients in stable and normal cases

Portfolio	Period, days	$c_{stable}^{opt} = b_0 + b_1 \cdot c_{normal}^{opt} + \xi_2$		
		b_0	b_1	$p\text{-level}$
DowJonesIndustrials and 3-monthT-bill	1563	0.0014	0.9907	0.0063
24 assets of DAX30	1387	0.0332	0.8730	0.0258
DAX30 and 6-month EURIBOR	1139	0.0022	0.9741	0.0057
DAX30, CAC40 and 6-month EURIBOR	1146	0.0351	0.8413	0.0104

the stable case is represented as a function of the corresponding risk aversion coefficient in normal case. The regression equation is the following:

$$c_{stable}^{opt} = b_0 + b_1 \cdot c_{normal}^{opt} + \xi_2. \quad (9)$$

The estimation of the regression coefficients in case of the DowJonesIndustrials index and 3-monthT-bill portfolio produces the following result:

$$c_{stable}^{opt} = 0.0014 + 0.9907 \cdot c_{normal}^{opt}, \quad (10)$$

where the $p\text{-level} = 0.0063$, which also speaks for the statistical significance of the estimated values and, therefore, of the prediction.

Table 2 presents the estimation of the regression coefficients for the other portfolios.

To summarize, the results indicate that the risk aversion coefficient value in the stable case is significantly lower than in the normal one, i.e. $c_{stable}^{opt} \ll c_{normal}^{opt}$. The difference becomes more apparent for higher values of the risk aversion coefficient. Thus, we can conclude that under the assumption that portfolio asset returns follow the stable distribution results in lower risk aversion coefficients compared with the case of the normal distribution.

58.8 Conclusions

In this chapter, we first briefly describe the stable distributions as an alternative approach to financial market investigation and its advantages compared with the normal distribution.

We propose a model for portfolio optimization under the assumption that portfolio asset returns are subject to the stable distribution. The comparison made between the stable and the normal approach in terms of the optimal allocation problem indicates that the stable distribution assumption for asset returns impacts the analysis in a meaningful way, as compared with the normal distribution. The stable approach considers that risk results from heavy tails. Thus the heavy tail behaviour of the stable distributed asset returns, compared with the normal case, greatly affects decision making in asset allocation. Since the stable distribution reflects better the financial market reality, better results in portfolio optimisation obtain under the stable distribution assumption.

Finally, we investigate the relationship between the maximum expected utility functions of an investor in stable and normal cases. We analyze the optimal risk aversion coefficients ratio obtained in both normal and stable cases. The analysis highlights significant differences in the stable and normal approaches, where the assumption of the stable distributed asset returns allowed for improvement on performance of the model.

References

- Abel, A.B. 1990. Asset prices under habit formation and catching up with the Joneses. *A.E.R. Papers and Proceedings* 80: 38–42.
- Aiyagari, S.R., and M. Gertler. 1991. Asset returns with transactions costs and uninsured individual risk. *Journal of Monetary Economics* 27: 311–331.
- Alvarez, F., and U. Jermann. 2000. Asset pricing when risk sharing is limited by default. *Econometrica* 48: 775–797.
- Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9: 203–228.
- Attanasio, O.P., J. Banks, and S. Tanner. 2002. Asset holding and consumption volatility. *Journal of Political Economy* 110: 771–792.
- Baba, N. 2000. *Testing the ex-ante relationship between asset and investment returns in Japan: An application of the P-CAPM to the Japanese asset returns*. IMES discussion paper series 2000-E-5.
- Balke, N.S., and T.B. Fomby. 1994. Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series. *Journal of Applied Econometrics* 9: 181–200.
- Bansal, R., and J.W. Coleman. 1996. A monetary explanation of the equity premium, term premium and risk free rate puzzles. *Journal of Political Economy* 104: 1135–1171.
- Bansal, R., and A. Yaron. 2000. *Risks for the long run: A potential resolution of asset pricing puzzles*. Working paper 8059 (NBER).
- Barberis, N., M. Huang, and T. Santos. 2001. Prospect theory and asset prices. *Quarterly Journal of Economics* 116: 1–53.
- Basak, S., and D. Cuoco. 1998. An equilibrium model with restricted stock market participation. *Review of Financial Studies* 11: 309–341.
- Basak, S., and A. Shapiro. 2001. Value-at-risk based risk management: Optimal policies and asset prices. *Review of Financial Studies* 14: 371–405.
- Benartzi, S., and R.H. Thaler. 1995. Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics* 110: 73–92.
- Bewley, T.F. 1982. *Thoughts on tests of the intertemporal asset pricing model*. Manuscript. Evanston: Northwestern University.

- Boldrin, M.L., L.J. Christiano, and J.D.M. Fisher. 2001. Habit persistence, asset returns, and the business cycle. *American Economic Review* 91: 149–166.
- Bossaerts, P. 1995. The econometrics of learning in financial markets. *Econometric Theory* 11: 151–189.
- Brav, A., G.M. Constantinides, and C.C. Geczy. 2002. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy* 110: 793–824.
- Brown, S., W. Goetzmann, and S. Ross. 1995. Survival. *Journal of Finance* 50: 853–873.
- Campbell, J.Y. 2001. Asset pricing at the millennium. *Journal of Finance* 55: 1515–1567.
- Campbell, J.Y., and J.H. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107: 205–251.
- Clark, P. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41: 135–155.
- Constantinides, G.M. 1990. Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy* 98: 519–543.
- Constantinides, G.M., and D. Duffie. 1996. Asset pricing with heterogeneous consumers. *Journal of Political Economy* 104: 219–240.
- Constantinides, G., J. Donaldson, and R. Mehra. 1998. *Junior can't borrow: A new perspective on the equity premium puzzle*, NBER working paper series no. 6617. Cambridge, MA: National Bureau of Economic Research.
- Constantinides, G.M., J.B. Donaldson, and R. Mehra. 2002. Junior can't borrow: A new perspective on the equity premium puzzle. *Quarterly Journal of Economics* 118: 269–296.
- Daniel, K., and D. Marshall. 1997. The equity premium puzzle and the risk-free rate puzzle at long horizons. *Macroeconomic Dynamics* 1: 452–484.
- Danthine, J.-P., J.B. Donaldson, and R. Mehra. 1992. The equity premium and the allocation of income risk. *Journal of Economic Dynamics and Control* 16: 509–532.
- Duffie, D., and J. Pan. 1997. An overview of value at risk. *Journal of Derivatives* 4: 7–49.
- Engsted, T., E. Mammen, and C. Tanggaard. 2001. *Evaluating the C-CAPM and the equity premium puzzle at short and long horizons: A Markovian bootstrap approach*. Aarhus: Mimeo, Aarhus School of Business.
- Epstein, L.G., and S.E. Zin. 1991. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy* 99: 263–286.
- Fama, E. 1965. Portfolio analysis in a stable Paretian market. *Management Science* 11: 404–419.
- Ferson, W.E., and G.M. Constantinides. 1991. Habit persistence and durability in aggregate consumption. *Journal of Financial Economics* 29: 199–240.
- Gabaix, X., and D. Laibson. 2001. The 6D bias and the equity premium puzzle. In *NBER macroeconomics annual*, ed. B. Bernanke and K. Rogoff. Cambridge, MA: MIT Press.
- Hansen, L.P., and R. Jagannathan. 1991. Implications of security market data for models of dynamic economies. *Journal of Political Economy* 99: 225–262.
- He, H., and D.M. Modest. 1995. Market frictions and consumption-based asset pricing. *Journal of Political Economy* 103: 94–117.
- Heaton, J. 1995. An empirical investigation of asset pricing with temporally dependent preference specifications. *Econometrica* 66: 681–717.
- Heaton, J., and D.J. Lucas. 1996. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy* 104: 443–487.
- . 1997. Market frictions, savings behavior and portfolio choice. *Journal of Macroeconomic Dynamics* 1: 76–101.
- Huisman, R., K.G. Koedijk, and R.A.J. Pownall. 1999. *Asset allocation in a value-at-risk framework*, Working paper. Rotterdam: Erasmus University of Rotterdam.
- İmrohoroglu, S. 2003. *A note on the McGrattan and Prescott (2003) adjustments and the equity premium puzzle*, Working paper. Los Angeles: University of Southern California.
- Jacobs, K. 1999. Incomplete markets and security prices: Do asset-pricing puzzles result from aggregation problems? *Journal of Finance* 54: 123–163.

- Janicki, A., and A. Weron. 1994. *Simulation and chaotic behaviour of stable stochastic processes*. New York: Marcel Dekker.
- Kocherlakota, N.R. 1996. The equity premium: It's still a puzzle. *Journal of Economic Literature* 34: 42–71.
- Krebs, T. 2000. *Consumption-based asset pricing with incomplete markets*, Working paper. Providence: Brown University.
- Kurz, M., and A. Beltratti. 1996. *The equity premium is no puzzle*. Stanford economics working paper #96-004.
- Kydland, F., and E.C. Prescott. 1982. Time to build and aggregate fluctuations. *Econometrica* 50: 1345–1371.
- Lettau, M. 2001. *Idiosyncratic risk and volatility bounds, or can models with idiosyncratic risk solve the equity premium puzzle?* New York: Federal Reserve Bank of New York and CEPR.
- Lucas, D.J. 1994. Asset pricing with undiversifiable risk and short sales constraints: Deepening the equity premium puzzle. *Journal of Monetary Economics* 34: 325–341.
- Lucas, R.E., Jr. 1978. Asset prices in an exchange economy. *Econometrica* 46: 1429–1445.
- Luttmer, E.G.J. 1996. Asset pricing in economies with frictions. *Econometrica* 64: 1439–1467.
- Lynch, A.W. 1996. Decision frequency and synchronization across agents: Implications for aggregate consumption and equity returns. *Journal of Finance* 51: 1479–1497.
- Macaulay, F.R. 1938. *The movements of interest rates, bond yields and stock prices in the United States since 1856*. New York: National Bureau of Economic Research.
- Mandelbrot, B. 1963a. The variation of certain speculative prices. *Journal of Business* 36: 394–419.
- . 1963b. The stable Paretian income distribution when the apparent exponent is near two. *International Economic Review* 4: 111–114.
- Mandelbrot, B., and M. Taylor. 1967. On the distribution of stock price differences. In *Operations research no. 15*, pp. 1057–1062.
- Mankiw, N.G. 1986. The equity premium and the concentration of aggregate shocks. *Journal of Financial Economics* 17: 211–219.
- Mankiw, N.G., and S.P. Zeldes. 1991. The consumption of stockholders and nonstockholders. *Journal of Financial Economics* 29: 113–135.
- McGrattan, E.R., and E.C. Prescott. 2001. *Taxes, regulations, and asset prices*, Working paper 610. Minneapolis: Federal Reserve Bank of Minneapolis.
- . 2003. Average debt and equity returns: Puzzling? *American Economic Review* 93: 392–397.
- Mehra, R., and E.C. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15: 145–161.
- . 1988. The equity premium: A solution? *Journal of Monetary Economics* 22: 133–136.
- . 2003. The equity premium in retrospect. In *Handbook of the economics of finance*, 887–936. New York: Elsevier.
- Melino, A., and A.X. Yang. 2003. *State dependent preferences can explain the equity premium puzzle*, Working paper. Toronto: University of Toronto.
- Merton, R.C. 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8: 323–361.
- Mittnik, S., S. Rachev, and M. Paoletta. 1997. Stable Paretian modelling in finance: Some empirical and theoretical aspects. In *A practical guide to heavy tails: Statistical techniques for analyzing heavy tailed distributions*, ed. R.J. Adler et al. Boston: Birkhauser.
- Officer, R.R. 1972. The distribution of stock returns. *Journal of the American Statistical Association* 67: 807–812.
- Ortobelli, S., I. Huber, and E. Schwartz. 2002. Portfolio selection with stable distributed returns. In *Mathematical methods of operations research*, 265–300. Cham: Springer.
- Ortobelli, S., I. Huber, S. Rachev, and E. Schwartz. 2003. Chapter 14. Portfolio choice theory with non-Gaussian distributed returns. In *Handbook of heavy tailed distributions in finance*, North Holland handbooks of finance, series editor W. T. Ziemba, 547–594. Amsterdam: North Holland.

- Otrok, Ch., B.Ravikumar, and Ch.H.Whiteman. 2001. Stochastic discount factor models and the equity premium puzzle.
- Rietz, T.A. 1988. The equity risk premium: A solution. *Journal of Monetary Economics* 22: 117–131.
- Samorodnitsky, G., and M. Taqqu. 1994. *Stable non-Gaussian random processes: Stochastic models with infinite variance*. New York: Chapman and Hall.
- Storesletten, K., C.I. Telmer, and A. Yaron. 2001. *Asset pricing with idiosyncratic risk and overlapping generations*, Working paper. Pittsburgh: Carnegie Mellon University.
- Teichmoeller, J. 1971. A note on distribution of stock price changes. *Journal of the American Statistical Association* 66: 282–284.
- Telmer, C.I. 1993. Asset-pricing puzzles and incomplete markets. *Journal of Finance* 49: 1803–1832.
- Vissing-Jorgensen, A. 2002. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy* 110: 825–853.
- Weil, P. 1989. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics* 24: 401–421.
- Ziemba, W.T. 1974. Choosing investments when the returns have stable distributions. In *Mathematical programming in theory and practice*, ed. P.L. Hammer and G. Zoutendijk, 443–482. Amsterdam: North-Holland.



Understanding Ginnie Mae Reverse Mortgage H-REMICs: Its Programs and Cashflow Analysis

59

C. H. Ted Hong and George H. Lee

Abstract

A reverse mortgage is a financial product that allows a senior homeowner to access their home equity. In contrast to a traditional or conventional “forward” mortgage, a reverse mortgage allows the borrower to draw on the equity in the home and only repay the loan when they vacate the house. This allows a borrower with few liquid assets but a lot of accumulated home equity to access trapped cash. The FHA’s reverse mortgage program referred to as the Home Equity Conversion (HECM) program was established in 1989, though Ginnie Mae mortgage-backed securities backed by HECM’s (HMBS) were issued for investors and the first Ginnie Mae H-REMIC (CMO) backed by multiple HMBS pools was not issued until 2009. The methods and parameters used to compute yields with varying LIBOR assumptions for CMO tranches backed by reverse mortgages differs from those for traditional mortgages.

Keywords

HECM · HMBS · H-REMIC · Reverse mortgage

The Author would like to thank Rico Fu for preparing the H-REMIC Analytics in the Beyondbond, Inc. proprietary systems and Joseph J. Kelly from New View Advisors LLC for his invaluable assistance in validating our computational results.

C. H. Ted Hong (✉) · G. H. Lee
Beyondbond, Inc, New York, NY, USA
e-mail: ted@beyondbond.com

59.1 Introduction

A reverse mortgage is a type of mortgage that borrowers, instead of paying the lender, will receive fixed or variable cash flow payments before the maturity date or when the home owner vacates the property used as collateral. The loan is only repaid if the property is sold or refinanced. A reverse mortgage is designed for senior home owners, 62 years of age or older, in the U.S. to maximize the duration of receiving payments against their properties. As baby boomers started to retire and the demographic pyramid gradually flattened, the eligible pool of borrowers for reverse mortgages increased in the past decade (see NRMLA Website). There are three types of reverse mortgages:

1. Single purpose reverse mortgages – these are offered only by some state and local government agencies and nonprofit organizations, like credit unions.
2. FHA insured reverse mortgages – also known as Home Equity Conversion Mortgages (HECMs) and backed by the US Department of Housing and Urban Development (HUD).
3. Private/Proprietary reverse mortgages – private loans, typically backed by the insurance or specialty finance companies that originate them.

Single purpose reverse mortgages are not available everywhere and can only be used for the purpose specified by the lender or the government, for example, a loan is made for repairs or property taxes only. The most popular type for U.S. seniors is the HECM program. The HECM is run by the U.S. Department of Housing and Urban Development through the Federal Housing Administration (FHA).

59.2 The FHA Home Equity Conversion Mortgage (HECM) Program

The FHA's HECM platform has a dominant market share in excess of 90% and is where we focus further analysis and discussion. The FHA's Home Equity Conversion Mortgage (HECM) program, which was piloted in 1989 and made permanent in 1998, allows older homeowners to withdraw some of the equity in their homes.

In addition, a HECM mortgage may be used to purchase a primary home when the borrower is 62 years of age or older and is able to use assets to fund the difference between the reverse mortgage and the sales price plus closing costs.

Unlike for a traditional forward mortgage, the HECM borrower does not make payments to the lender; rather, the lender makes payments to the borrower. A HECM mortgage allows a borrower to extract a portion of the equity in the home while continuing to live in the residence, and in that sense, a HECM resembles a Home Equity Line of Credit (HELOC). HECM loan origination has grown steadily since the program's inception in 1989. Figure 1 shows that annual loan origination has increased from \$544 million in 2000, peaking at \$21.7 billion in 2009, to \$10.9

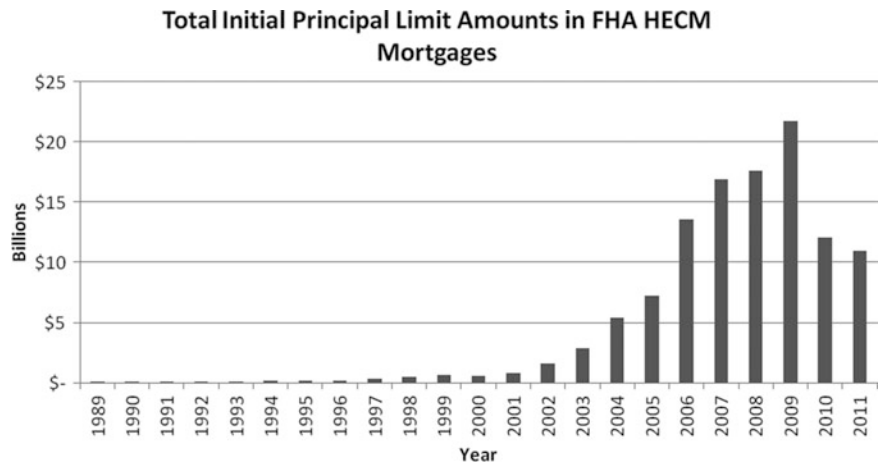


Fig. 1 FHA origination volume by initial principal limit. (Source: FHA, Beyondbond, Inc.)

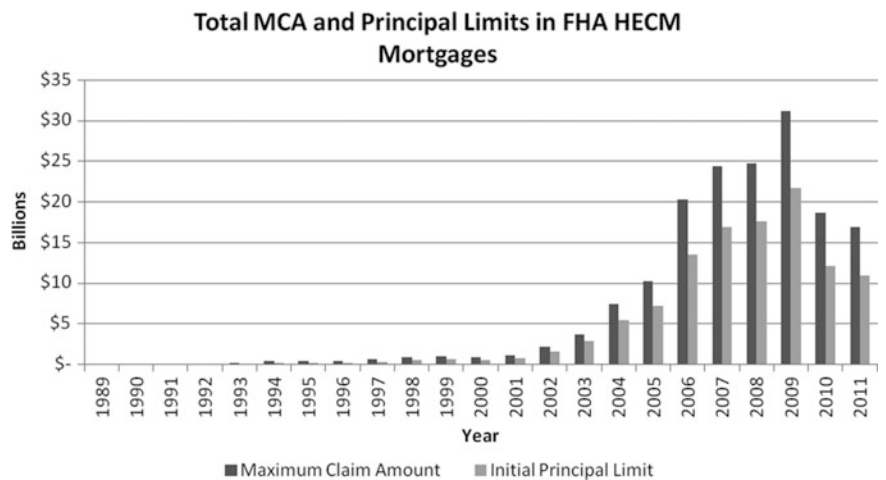


Fig. 2 FHA origination volume by maximum claim amount and initial principal limit. (Source: FHA, Beyondbond, Inc.)

billion for 2011. Figure 2 adds the maximum claim about to the initial principal limit to the annual loan origination chart.

59.2.1 HECM Versus HELOC

A HECM mortgage allows a borrower to extract a portion of the equity in the home, while continuing to live in the residence, and in that sense, a HECM resembles a

traditional or conventional Home Equity Line of Credit (HELOC). However, a HECM differs from a HELOC in the following ways:

- First, no principal repayment is required until the borrower no longer uses the home as a primary residence. Unlike a traditional HELOC, the payments are reversed, so the principal payments flow from the lender to the borrower.
- Second, the borrower is required to use the house as the primary residence. Per FHA underwriting standards, the residence must be owner occupied.
- Third, a HELOC requires a sufficient debt-to-income ratio to service the loan and a combined LTV test as well as FICO score. Since the payments are reversed, lenders do not consider the borrower's debt to income ratio for a HECM or require a FICO.

59.2.2 Qualifications and Payment Plans

HUD has listed the FHA qualified borrower requirements and various payment plans for HECM summarized as the following table¹ (see Ginnie Mae MBS Guide 2011; HUD Handbook 1994):

FHA Eligibility requirements	Borrow requirements	Payment plans
The borrower must be 62 years of age or older	The borrower must be 62 years of age or older	The Tenure option provides for equal monthly payments as long as at least one borrower lives and continues to occupy the home as his/her principal residence
Ownership of the property must be outright or a de minimus mortgage balance	Ownership of the property must be outright or a de minimus mortgage balance	
The home must be the principal residence	The home must be the principal residence	The Term option provides for equal monthly payments for a fixed period of months selected by the borrower
The borrower may not be delinquent on any federal debt	The borrower may not be delinquent on any federal debt	
Finally, counseling is mandatory. The borrower must participate in a consumer information session given by an approved HECM counselor	Finally, counseling is mandatory. The borrower must participate in a consumer information session given by an approved HECM counselor	The Line of Credit option allows the borrower to draw (unscheduled payments) at times and in an amount of his/her choosing until the line of credit is exhausted
HUD-approved properties include single-family or 1–4 unit homes, condominiums and manufactured homes	HUD-approved properties include single-family or 1–4 unit homes, condominiums and manufactured homes	The Modified Tenure option is a combination of scheduled monthly payments as long as the

(continued)

¹See Department of Housing and Urban Development website with respect to “FHA Reverse Mortgages (HECMs) for Seniors” section.

FHA Eligibility requirements	Borrow requirements	Payment plans
		borrower remains in the home and a line of credit
		The Modified Term option is a combination of scheduled monthly payments for a fixed time period and a line of credit

59.2.3 Loan Amount: Principal Limit; Maximum Claim Amount (MCA)

The amount of money the homeowner can borrow is known as the principal limit, and is calculated using a formula based on borrower’s age, current interest rate level and home value (see Ginnie Mae MBS Guide 2011). As older borrowers are expected to accrue less interest, and therefore may borrow more. The lower interest rate environment obviously will lead to less interest accrued, and allow the homeowner to borrow more. The FHA has county-based lending limits based on borrower’s location and home value. The amount available is based on the lesser of the home value and the FHA limit for the county. The national lending limit is \$625,500.

It is important to note that the principal limit is less than the Maximum Claim Amount (MCA). The maximum claim amount is the largest amount that the FHA will insure for a given property, and is equal to the lesser of the home value and the local FHA lending limit. Since interest is capitalized, the principal can grow above the principal limit. However, when the principal nears the MCA, the FHA guarantee and put take effect.

59.2.4 Fees and Caps

FHA caps many of the up-front fees and rates associated with the HECM. For example: (1) Mortgage insurance premium (MIP): The borrower pays a MIP equal to 2% of the maximum claim amount (the value of the home or FHA county lending limit, whichever is less), plus an annual premium thereafter equal to 0.5% of the loan balance. The MIP is paid directly to FHA in exchange for guaranteeing the loan. The MIP guarantees that if the company managing the Borrower’s account, either the servicer or issuer goes out of business, the government will step in and make sure the borrower has continued access to the borrower’s loan funds. (2) Rate adjustment cap: In the case of an annually adjusting loan, the rate cannot increase more than 2% per year. However, in the case of a monthly adjusting loan, there is no annual cap. (3) Lifetime rate adjustment cap: In the

case of an annually adjusting loan, the rate adjustment is capped at 5% over the initial loan rate at closing. In the case of a monthly adjusting loan, the cap is 10% over the initial loan rate at closing. (4) Servicing fees: Fees are capped at \$30 per month for annually adjusted loans; \$35 per month for monthly adjusted loans. A present value estimate of future service fees is set aside at closing based on the assumption the homeowner will live until the age of 100. (5) There is also an origination fee charged on every loan. These vary from lender to lender. Lenders usually allow borrowers to finance the origination fee, and the FHA caps the fee at the greater of \$2000 or 2% of the MCA.

In addition to the fees above, the borrower will also pay other standard closing costs associated with getting a mortgage, including title insurance, attorney's fees, recording taxes, etc.

59.2.5 HECM Prepayment Curve (PPC)

The pricing convention in the HECM market is based on a prepayment curve (PPC vector). The HECM PPC ramp begins at 0.05% CPR and increases to 12% CPR in month 39. The PPC ramp continues to increase through month 360, reaching at peak at 43% CPR. The PPC ramp is also referred to as the FHA or HUD curve and is based on the prepayment experience of FHA HECM loans and is shown graphically in Fig. 3 and complete details are included in Appendix 2.

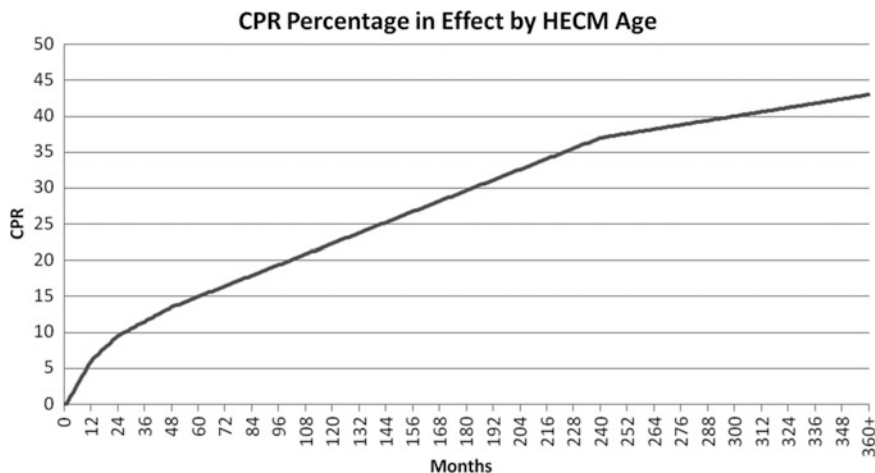


Fig. 3 The HECM PPC by age. (Source: FHA, Ginnie Mae)

59.2.5.1 Ginnie Mae HECM Pooling (HMBS)

The Government National Mortgage Association (Ginnie Mae) has created a HECM loan pooling program (HMBS) with the intent on providing liquidity to the reverse mortgage market. HECM MBS are backed by participation interests in advances made to borrowers and related amounts. As Ginnie Mae securities, HECM's share the following characteristics with other Ginnie Mae programs. First, Ginnie Mae reverse mortgage loans are insured by the Federal Housing Administration (FHA), and Ginnie Mae guarantees the timely payment of principal and interest on each class of securities. Second, the Ginnie Mae guarantee is backed by the full faith and credit of the United States. Third, HECMs are Ginnie Mae II securities and, as such, carry a 0% risk weighting for capital purposes.

HECM balances are funded at the time the borrower draws on his/her line of credit. Because the borrower may not draw the entire balance available at the time of loan closing, a single borrower may be represented in several HMBS pools or securitizations.

The issuer/servicer is also obligated to buy out any advance from a pool when the underlying loan reaches 98% of its MCA. This flows through to the investor as a prepayment, not as scheduled P&I. If the loan has become due, the servicer will continue to service the loan until the home is sold and the loan is liquidated. If there is any shortfall, the FHA guarantee will come into effect and the FHA will make the servicer whole.

If the loan has not become due, the loan can be assigned to the FHA, and the servicer is again made whole. This typically happens when a borrower stays in the home longer than expected. This is also referred to as the FHA put.

59.2.5.2 HECM Pool Characteristics

Similar to traditional forward mortgages, reverse mortgage pools will have aggregation-level data reported on them. These will include the (1) Single male, single female and couple ages, (2) Loan and property type, (3) Principal limit, (4) Ratio of outstanding balance to MCA, and (5) Ratio of outstanding balance to principal limit.

Similar to forward mortgage pass-throughs, there is both a note rate (rate paid by the borrower) and a pass-through rate (rate received and accrued to the security holder). The spread, which compensates both the issuer and Ginnie Mae, can range from 6 to 75 bps (see Rao 2009; Ginnie Mae MBS Guide 2011).

Ginnie Mae mandates that a pool be at least \$1MM in original balance, and contain at least three HECM participations from three distinct HECM reverse mortgages.

59.2.5.3 Cash Flows and Prepayments

A reverse mortgage advance is similar to a zero-coupon bond to an investor, but has an uncertain maturity. The investor purchases the pool, and interest accrues until the

borrower chooses to prepay, or the loan comes due. Borrowers are permitted to prepay their reverse mortgages at any time without penalty.

When a loan comes due, the borrower (or their estate) has 12 months to either sell the home or pay off the reverse mortgage with cash. The loan continues to accrue interest during this time.

Investors also get a prepayment when the principal balance on the HECM is equal to or greater than 98% of the MCA. At that point, as mentioned previously, the issuer is obligated to buy out all participations from that HECM (most likely in multiple securities). Since the accrued interest is added on to the loan principal, any mortgage will eventually reach this limit.

Historically, 90% of reverse mortgages are paid off through the sale of the home, refinancing, or with cash (see Patterson et al. 2007). 9% of loans are bought out by the lender and assigned to the FHA. The remaining 1% requires the FHA to step in and pay a claim because the proceeds of the home sale were not enough to cover the loan.

Also, it is important to note that because interest is capitalized into the loan, every HECM will eventually be bought out by the issuer even if there is no maturity event. According to HUD's data, 58% of HECMs terminate or are assigned within 7 years, and 100% of HECMs terminate or are assigned within 18 years. Thus, even though a HMBS may have a stated final maturity of 50 years, it is very unlikely that any participation will remain in the pool after even half that time.

59.2.5.4 H-REMIC Tranche Principal Types

HMBS can be further aggregated and structured in the form of a Reverse Mortgage CMO or H-REMIC. An H-REMIC provides for additional diversification as well as specific tranching of cashflows for specific investor needs (see Schultz and Ahlgren 2010). Figure 4 shows similar growth in the H-REMIC monthly issuance since inception. As described in Fig. 5, there are currently three principal bond types in an H-REMIC:

- **HECM MBS Pass-Through Class:** A HECM MBS Pass-Through Class ("HPT") is a Class that is backed in whole or in part by Trust Assets consisting of HECM MBS that either individually or together with other Classes receives on each Distribution Date all, or substantially all, of the principal and/or interest payments received on the related Trust Assets and that is not a Strip or HECM MBS Sequential Pay Class.
- **HECM MBS Sequential Pay Class:** A HECM MBS Sequential Pay Class ("HSEQ") is a Class that is backed in whole or in part by Trust Assets consisting of HECM MBS that receives distributions of principal in a prescribed sequence, that do not have predetermined schedules and that generally are designed to receive distributions of principal continuously from the first Distribution Date on which they receive principal until they are retired. A HECM MBS Sequential Pay Class may receive principal distributions concurrently with one or more other HECM MBS Sequential Pay Classes.

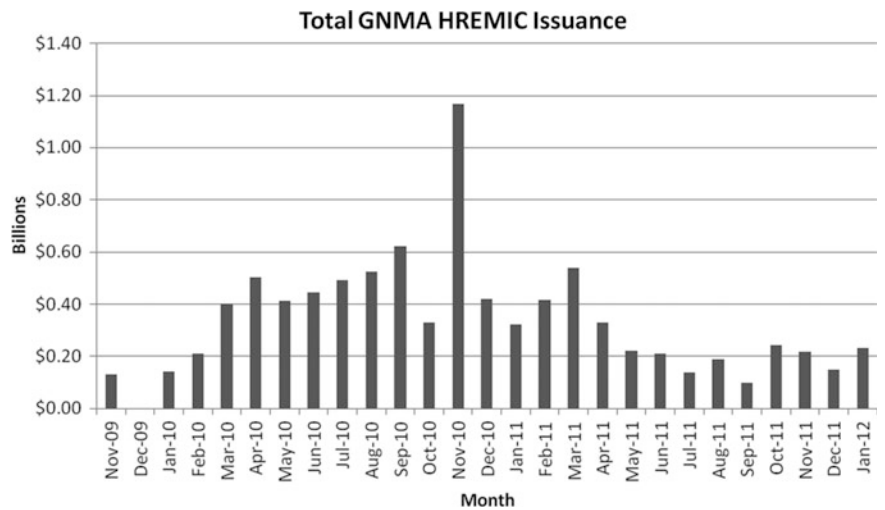


Fig. 4 Ginnie Mae H-REMIC issuance. (Source: Ginnie Mae, Beyondbond, Inc.)

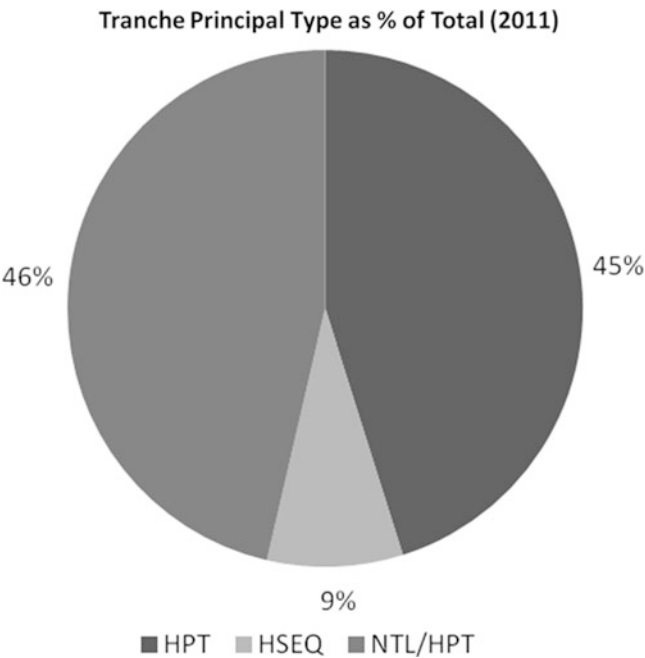


Fig. 5 Ginnie Mae H-REMIC tranche principal types. (Source: Ginnie Mae, Beyondbond, Inc.)

- Notional Classes: The Notional Classes (“NTL”) will not receive distributions of principal based on their Class Notional Balances but have Class Notional Balances for convenience in describing their entitlements to interest.

59.2.6 H-REMIC Modeling Example: GNR 2011-H10

We will analyze a Ginnie Mae Reverse Mortgage CMO/H-REMIC, GNR 2011-H10 (see Ginnie Mae REMIC Offering Circulars 2010), using the Beyondbond, Inc. proprietary reverse mortgage cashflow modeling analytic system.

The Issuer (Ginnie Mae), Dealer and Accountant must tie out all aspects of the computations contained in the Prospectus Supplement, which includes Weighted Average Life, Decrement (DEC) Tables, and Price-Yield Tables.

Because of the complexity in the modeling of cashflows, certain standard modeling assumptions are made.

59.2.6.1 Assumptions for GNR 2011-H10

The tables that follow have been prepared on the basis of the following assumptions (the “Modeling Assumptions”), among others:

1. The HECMs and related Participations underlying the Trust Assets have the assumed characteristics shown in the Beyondbond, Inc. Collateral Viewer in Fig. 6 and further detailed in Appendix 1.
2. The HECMs prepay at the constant percentages of the prepayment curve (shown in Fig. 3) shown in the related table in Appendix 2.
3. Draw activity occurs on the first day of the month and payments on the HECMs occur on the last day of the month, whether or not a Business Day, commencing in April 2011.
4. Distributions, if any, on the Securities are always received on the 20th day of the month, whether or not a Business Day, commencing in May 2011.
5. A termination of the Trust does not occur.
6. The Closing Date for the Securities is April 29, 2011.
7. No expenses or fees are paid by the Trust other than the Trustee Fee.
8. HECM borrowers who have the ability to do so draw at the annualized draw rate determined in accordance with the constant percentages of the draw curve shown in Exhibit C (the “Draw Rate”). The Draw Rate (converted to an equivalent monthly factor) is applied to the Maximum Claim Amount.
9. If a mandatory Ginnie Mae Issuer Purchase Event occurs with respect to a HECM, the purchase of the related Participation timely occurs. No optional Ginnie Mae Issuer Purchase Events occur.
10. The initial value of LIBOR on the Securities is 0.26%; however, the interest rate on the adjustable rate HECMs for the first Distribution Date is based on the information set forth in Exhibit A. On all Distribution Dates occurring after the

Loan ID	Payment Plan	Int Type	HMBS Balance	HECM Balance	Loan Age	Gross Coupon %	Gross Margin %	Lifetime Floor %	Lifetime Cap %	MIP Fee %	Servicing Fee %	Index
1	Line_of_Credit	Floating	18,619,073.29	149,868,750.70	20.00	3.26	3.00	3.00	13.28	0.50	0.06	1-month-LIBOR
2	Modified_Tenure	Floating	907,305.03	4,392,716.09	19.00	3.26	3.00	3.00	13.28	0.50	0.06	1-month-LIBOR
3	Modified_Tenure	Floating	2,100,257.97	9,203,544.50	19.00	3.26	3.00	3.00	13.27	0.50	0.06	1-month-LIBOR
4	Tenure	Floating	542,880.48	3,936,940.86	19.00	3.26	3.00	3.00	13.27	0.50	0.06	1-month-LIBOR
5	Tenure	Floating	97,118.96	2,577,209.70	22.00	3.26	3.00	3.00	13.36	0.50	0.06	1-month-LIBOR
6	Line_of_Credit	Floating	26,378,756.79	28,212,847.90	31.00	1.14	0.89	0.89	13.37	0.50	0.09	1-month-LIBOR
7	Modified_Tenure	Floating	800,558.63	1,039,739.89	32.00	1.23	0.97	0.97	13.45	0.50	0.09	1-month-LIBOR
8	Modified_Tenure	Floating	1,217,565.63	1,606,802.18	32.00	1.11	0.85	0.85	13.33	0.50	0.09	1-month-LIBOR
9	Tenure	Floating	225,384.11	278,675.63	32.00	1.08	0.82	0.82	13.29	0.50	0.09	1-month-LIBOR
10	Tenure	Floating	159,213.20	202,880.29	31	1.241	0.982	0.982	13.462	0.5	0.09	1-month-LIBOR
11	Line_of_Credit	Floating	36,059,285.68	38,263,386.61	30	1.681	1.421	1.421	13.942	0.5	0.09	1-month-LIBOR
12	Modified_Tenure	Floating	404,228.16	584,857.30	30	1.759	1.499	1.499	14.477	0.5	0.09	1-month-LIBOR
13	Modified_Tenure	Floating	500,822.39	637,769.39	31	1.679	1.419	1.419	13.799	0.5	0.09	1-month-LIBOR
14	Tenure	Floating	262,154.81	304,121.40	29	1.65	1.39	1.39	13.183	0.5	0.09	1-month-LIBOR
15	Tenure	Floating	203,685.31	242,586.98	31	1.724	1.464	1.464	13.698	0.5	0.09	1-month-LIBOR
16	Line_of_Credit	Floating	43,771,754.23	47,178,372.92	30	1.409	1.156	1.156	14.532	0.5	0.09	1-month-LIBOR
17	Modified_Tenure	Floating	1,420,453.11	1,796,027.09	31	1.421	1.164	1.164	14.38	0.5	0.09	1-month-LIBOR
18	Modified_Tenure	Floating	1,753,709.43	2,426,214.92	30	1.43	1.173	1.173	14.518	0.5	0.09	1-month-LIBOR
19	Tenure	Floating	336,097.61	443,640.54	31	1.43	1.173	1.173	14.6	0.5	0.09	1-month-LIBOR

Fig. 6 GNR 2011-H10 collateral viewer. (Source: Beyondbond, Inc.)

first Distribution Date, the value of LIBOR on the adjustable rate HECMs is assumed to be the same as the value of LIBOR on the Securities. For purposes of the decrement tables, on all Distributions Dates occurring after the first Distribution Date the constant value of LIBOR shown with respect to any decrement table is used to calculate the interest rate with respect to the adjustable rate HECMs and to the applicable Class.

11. The original term of the HECMs is 50 years.
12. No borrower changes payment plans.
13. Draws occur each month in respect of the Monthly Servicing Fee, if any, as set forth on Exhibit A. No draws occur in respect of any set asides for property charges (such as taxes, hazard insurance, ground rents or assessments) or repairs.

When reading the tables and the related text, investors should bear in mind that the Modeling Assumptions, like any other stated assumptions, are unlikely to be entirely consistent with actual experience.

For example, most of the HECMs will not have the characteristics assumed, many Distribution Dates will occur on a Business Day after the 20th of the month, draw activity and prepayments, if any, will occur throughout the month, draws will occur in respect of set asides for property charges and repairs, the Trustee may cause a termination of the Trust as described under “Description of the Securities – Termination” in this Supplement and LIBOR on the Class AF Securities may differ from LIBOR on the adjustable rate HECMs.

In addition, distributions on the Securities are based on Certificate Factors and Calculated Certificate Factors, if applicable, which may not reflect actual receipts on the Trust Assets.

59.2.7 Payment Rules and Tranche Waterfall

Each class in the prospectus is described as the following Fig. 7 for the REMIC class name, principal and interest types, annualized interest rate level (Interest Rate), initial certificate balance (Principal Balance), CUSIP, and schedule date of final payment received (Final Distribution Date)².

On each Distribution Date, the cashflow distribution of each REMIC class received from the HMBS certificates is referred to as the payment rules or tranche waterfall. The Available Distribution Amount will be allocated in the following order of priority:

Class of REMIC	Principal Balance	Interest Rate	Principal Type	Interest Type	CUSIP Number	Final Distribution Date
AF	225,965,045	Adjustable	HPT	FLT/HWAC/HZ	38375BLJ9	August 2060
IF	225,965,045	Adjustable	NTL(HPT)	HWAC/IO/DLY	38375BLK6	August 2060
RR	0	0	NPR	NPR	38375BLL4	August 2060

Fig. 7 GNR 2011-H10 deal viewer. (Source: Beyondbond, Inc.)

- 1. Concurrently, to AF and IF, pro rata based on their respective Interest Accrual Amounts, up to the Class AF Interest Accrual Amount and the Class IF Interest Accrual Amount for such Distribution Date
- 2. To AF, in reduction of its Class Principal Balance, up to the amount of the Class AF Principal Distribution Amount for such Distribution Date, until retired
- 3. To IF, until the Class IF Deferred Interest Amount is reduced to zero

59.2.7.1 Analysis

Cashflows are generated for each individual HMBS in various scenarios. Initially, absent any payoff events (0 PPC), no cashflows are paid to the H-REMIC classes and the interest is capitalized as displayed in the Cashflow Viewer in Fig. 8, compared with the Cashflow under prepayment assumption at 100 PPC.

HMBS has historically traded at premiums to par as mortgages have higher spreads than typical LIBOR bonds. Due to investors' preference for par floating rate bonds, H-REMICs are able to strip off the spread premium by tranching an

0 PPC

Period	Remaining Balance	Gross Cashflow	Draw	Schedule Draw	Servicing Fee	MIP	LOC	MCA	Bond Int	Bond Principal	Bond Balance	Gross Payment
0	149,868,750.70	0	0	0	0	0	0	42,160,357.74	289,116,200.00	0	18,619,073.29	0.00
1	151,248,565.51	409,528.25	0	877,845.48	0	29,630.00	62,811.08	41,411,864.13	289,116,200.00	49,650.86	18,668,724.15	1,379,814.81
2	152,598,017.44	413,182.81	0	843,267.52	0	29,630.00	63,371.60	40,695,713.52	289,116,200.00	49,783.26	18,718,507.42	1,349,451.93
3	153,917,057.06	416,755.01	0	808,735.13	0	29,630.00	63,939.48	40,011,953.42	289,116,200.00	49,916.02	18,768,421.44	1,319,018.62
4	155,205,634.65	420,244.71	0	774,248.17	0	29,630.00	64,454.71	39,360,652.26	289,116,200.00	50,049.13	18,818,472.57	1,288,577.59
5	156,463,700.23	423,651.78	0	739,806.52	0	29,630.00	64,977.27	38,741,857.72	289,116,200.00	50,182.59	18,868,655.16	1,258,065.57
6	157,691,203.51	426,976.08	0	705,410.07	0	29,630.00	65,487.13	38,155,628.52	289,116,200.00	50,316.41	18,918,971.57	1,227,503.28
7	158,888,093.94	430,217.48	0	671,058.68	0	29,630.00	65,984.28	37,602,021.50	289,116,200.00	50,450.59	18,969,422.16	1,196,890.44
8	160,054,320.69	433,375.83	0	636,752.23	0	29,630.00	66,468.69	37,081,093.78	289,116,200.00	50,585.13	19,020,007.29	1,166,226.75
9	161,189,832.63	436,451.00	0	602,490.60	0	29,630.00	66,940.34	36,592,902.80	289,116,200.00	50,720.02	19,070,727.31	1,135,511.94
10	162,294,578.36	439,442.86	0	568,273.67	0	29,630.00	67,399.21	36,137,506.30	289,116,200.00	50,855.27	19,121,582.58	1,104,745.74
11	163,368,506.20	442,351.25	0	534,101.31	0	29,630.00	67,845.28	35,714,962.31	289,116,200.00	50,990.89	19,172,573.47	1,073,927.84
12	164,411,564.18	445,176.04	0	499,973.40	0	29,630.00	68,278.53	35,325,329.23	289,116,200.00	51,126.86	19,223,700.33	1,043,057.97

100 PPC

Period	Ending Balance	Gross Interest	Gross Cashflow	Draw	Schedule Draw	Servicing Fee	MIP	LOC	MCA	Bond Int	Bond Principal	Bond Balance	Gross Payment
0	149,868,750.70	0	0	0	0	0	0	42,160,357.74	289,116,200.00	0	0	18,619,073.29	0.00
1	150,115,964.24	409,528.25	1,132,601.28	877,845.48	0	29,630.00	62,811.08	41,101,757.85	286,951,198.46	49,650.86	90,146.96	18,528,926.33	1,379,814.81
2	150,281,117.37	410,088.76	1,174,193.64	836,952.84	0	29,408.12	62,897.05	40,777,827.39	284,726,540.44	49,410.47	94,622.54	18,434,303.78	1,339,346.77
3	150,364,775.02	410,427.41	1,215,354.95	796,456.08	0	29,180.13	62,948.99	39,088,511.82	282,443,630.33	49,158.14	99,040.48	18,335,263.30	1,299,012.61
4	150,367,584.57	410,545.80	1,256,028.70	756,179.14	0	28,946.16	62,967.15	38,133,707.07	280,103,907.00	48,894.04	103,397.74	18,231,865.56	1,258,838.25
5	150,307,622.21	410,445.76	1,278,811.68	716,745.37	0	28,706.38	62,951.80	37,217,555.53	277,740,897.73	48,618.31	105,599.29	18,126,266.27	1,218,849.31
6	150,185,744.39	410,176.67	1,301,084.87	677,655.05	0	28,464.21	62,910.53	36,339,575.99	275,355,446.26	48,336.71	107,760.69	18,018,505.57	1,179,207.06
7	150,002,837.60	409,740.88	1,322,830.11	639,119.02	0	28,219.73	62,843.69	35,499,261.05	272,948,396.07	48,049.35	109,881.45	17,908,624.12	1,139,923.32
8	149,759,821.43	409,140.82	1,344,025.81	601,144.11	0	27,973.05	62,751.66	34,698,082.92	270,520,997.61	47,756.33	111,960.57	17,796,661.55	1,101,009.64
9	149,457,647.54	408,379.01	1,364,651.08	563,739.14	0	27,724.23	62,634.82	33,929,492.20	268,072,907.65	47,457.77	113,997.13	17,682,666.42	1,062,477.20
10	149,097,294.46	407,458.05	1,384,690.01	526,911.93	0	27,473.38	62,493.57	33,196,918.12	265,606,181.29	47,153.78	115,990.70	17,566,675.72	1,024,136.93
11	148,679,770.61	406,380.64	1,404,123.26	490,669.87	0	27,220.58	62,328.32	32,503,770.34	263,121,278.97	46,844.47	117,940.39	17,444,735.33	986,599.41
12	148,206,113.22	405,149.51	1,422,932.27	455,019.96	0	26,965.92	62,139.50	31,843,439.78	260,619,065.86	46,529.96	119,845.34	17,328,889.98	949,274.89

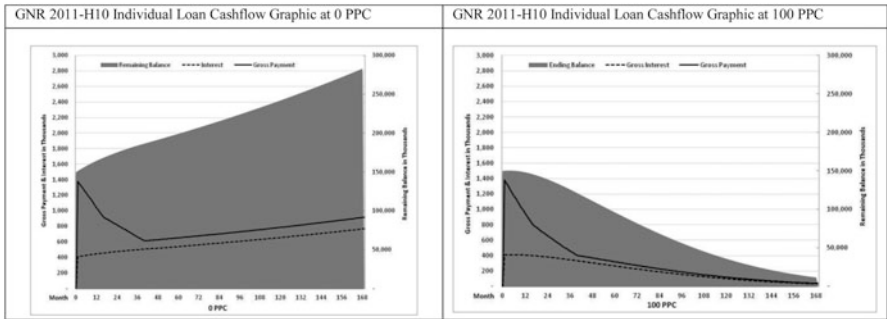


Fig. 8 GNR 2011-H10 individual loan cashflow viewer with prepayment assumption at 0 and 100 PPC. (Source: Beyondbond, Inc.)

LIBOR	0% PPC	75% PPC	100% PPC	125% PPC	150% PPC
0.26000%	0.72	0.71	0.71	0.71	0.70
1.26000%	1.72	1.71	1.70	1.70	1.69
5.52933%	6.01	5.99	5.98	5.97	5.96
9.79867%	10.33	10.30	10.29	10.28	10.26
14.06800%	11.60	11.56	11.55	11.53	11.52

Fig. 9 GNR 2011-H10 Class AF yield table. (Source: Beyondbond, Inc.)

LIBOR	0% PPC	75% PPC	100% PPC	125% PPC	150% PPC
0.26000%	9.43	11.28	6.96	2.61	-1.75
1.26000%	11.12	11.81	7.57	3.30	-1.01
5.52933%	13.68	10.15	6.44	2.67	-1.18
9.79867%	12.48	5.08	1.28	-2.17	-5.58
14.06800%	16.81	12.26	9.17	5.48	1.90

Fig. 10 GNR 2011-H10 Class IF yield table. (Source: Beyondbond, Inc.)

HMBS pool into a par floater, Class AF and an excess interest floater, Class IF. In the case of GNR 2011-H10, the underlying HMBS was trading at a premium of 105–106%. Class AF was issued at price of 100% and Class IF was then issued at a price of 5.5% (see Schultz 2010).

As a par floater pass-thru, Class AF shows stability in yields in Fig. 9, despite changes in prepayment rates and only variability due to changing LIBOR. This is desirable to investors as LIBOR can be readily hedged or directly moves in line with funding costs and is self-hedging.

In contrast to Class AF, which owes its yield stability to the fact that it receives both principal and interest, Class IF has a much more volatile yield pattern, as it only receives interest payments. Class IF could also be characterized as the premium component of the HMBS pass-thrus. Class IF appeals to an investor with a bearish view on repayments (bullish view on Borrowers longevity) or Administration healthcare initiatives (see Alexander 2008) (Fig. 10).

59.2.7.2 Conclusions and Further Study

Our analysis focuses on computing yields with varying HECM PPC and LIBOR assumptions for Ginnie Mae Reverse Mortgage H-REMIC (CMO) tranches as typically disclosed to investors in a Prospectus Supplement. Further study should explore evolving actuarial data to explore the influence on prepayment rates of

borrower age, gender, product type and geography for both FHA and Private loans.

In Appendix 4, we describe a way to estimate repayment rates from the existing US Life Tables (see Rai 2009) for mortality and mobility, however, Reverse Mortgages in some cases are being utilized in the Make Home Affordable (MHA) initiatives to help keep Senior Borrowers in their homes and avoid foreclosures (see Strand and Deb 2010). Any MHA or Administration initiative could structurally change the parameters for estimating prepayments.

59.2.8 Glossary (See Ginnie Mae MBS Guide 2011; NRMLA Website)

Reverse Mortgage	Reverse mortgage loans are FHA-insured loans designed specifically to permit senior citizens to convert the home equity of their principal residence into cash.
HECM	Home Equity Conversion Mortgages program. HECM was launched in 1987 by the Department of Housing and Urban Development. This is a federal reverse mortgage program and was launched to provide an opportunity to senior citizens to take advantage of their built-up home equity. The loans are fully guaranteed by the US government.
HMBS	Home Equity Conversion Mortgage Backed Security. It's collateralized by HECM loans.
H-REMIC	HECM Real Estate Mortgage Investment Conduit. Allows for inclusion of HMBS and forward Ginnie Mae MBS collateral within the same REMIC structure. First HREMIC Issuance: Anticipated January 2008.
HUD	Department of Housing and Urban Development. Congress created the HECM program in 1989 and appointed the Department of Housing and Urban Development (HUD) as the administrator.
MIP	Mortgage Insurance Premium. The borrower will be charged mortgage insurance premiums to reduce the risk of loss in the event that the outstanding balance, including accrued interest, MIP, and fees, exceeds the value of the property at the time that the mortgage is due and payable. HUD will select an agent to collect MIP.
MCA	Maximum Claim Amount. The amount that FHA will insure for any HECM loan. FHA allows Issuers to assign a HECM loan that accrues to 98% of the MCA. Ginnie Mae requires any loan that has accrued to 98% of MCA to be purchased out of an HMBS pool whether or not an Issuer assigns the loan to FHA.

Appendix 1: Assumed Trust Assets GNR 2011-H10: HMBS Pools (See Ginnie Mae REMIC Offering Circulars 2010)

Assumed characteristics of the HECMs and the participations underlying the trust assets(1)										
Payment plan	HECM MBS principal balance(2)	HECM loan balance	Approximate weighted average HECM age (in months) (3)	HECM interest type	Index	Rate reset frequency (4)	Next rate reset months (5)	Approximate weighted average gross interest rate (6)	Approximate weighted average gross margin(7)	approximate weighted average gross lifetime interest rate floor(8)
Line of credit	\$18,619,073.29	\$149,868,750.73	20	FLT	1-month LIBOR	Monthly	1	3.260%	3.000%	3.000%
Modified tenure	907,305.03	4,392,716.09	19	FLT	1-month LIBOR	Monthly	1	3.260	3.000	3.000
Modified term	2,100,257.97	9,203,544.50	19	FLT	1-month LIBOR	Monthly	1	3.260	3.000	3.000
Tenure	542,880.48	3,936,940.86	19	FLT	1-month LIBOR	Monthly	1	3.260	3.000	3.000
Term	97,118.96	2,577,209.70	22	FLT	1-month LIBOR	Monthly	1	3.260	3.000	3.000
Line of credit	26,378,756.79	28,212,847.90	31	FLT	1-month LIBOR	Monthly	1	1.144	0.886	0.886
Modified tenure	800,558.63	1,039,739.89	32	FLT	1-month LIBOR	Monthly	1	1.228	0.968	0.968
Modified term	1,217,565.63	1,606,802.18	32	FLT	1-month LIBOR	Monthly	1	1.112	0.853	0.853
Tenure	225,384.11	278,675.63	32	FLT	1-month LIBOR	Monthly	1	1.079	0.819	0.819
Term	159,213.20	202,880.29	31	FLT	1-month LIBOR	Monthly	1	1.241	0.982	0.982
Line of credit	36,059,285.68	38,263,386.61	30	FLT	1-month LIBOR	Monthly	1	1.681	1.421	1.421
Modified tenure	404,228.16	584,857.30	30	FLT	1-month LIBOR	Monthly	1	1.759	1.499	1.499
Modified term	500,822.39	637,769.39	31	FLT	1-month LIBOR	Monthly	1	1.679	1.419	1.419
Tenure	262,154.81	304,121.40	29	FLT	1-month LIBOR	Monthly	1	1.650	1.390	1.390
Term	203,685.31	242,586.98	31	FLT	1-month LIBOR	Monthly	1	1.724	1.464	1.464
Line of credit	43,771,754.23	47,178,372.92	30	FLT	1-month LIBOR	Monthly	1	1.409	1.156	1.156
Modified tenure	1,420,453.11	1,796,027.09	31	FLT	1-month LIBOR	Monthly	1	1.421	1.164	1.164
Modified term	1,753,709.43	2,426,214.92	30	FLT	1-month LIBOR	Monthly	1	1.430	1.173	1.173
Tenure	336,097.61	443,640.54	31	FLT	1-month LIBOR	Monthly	1	1.430	1.173	1.173
Term	209,903.41	243,198.71	30	FLT	1-month LIBOR	Monthly	1	1.430	1.173	1.173
Line of credit	779,906.80	18,243,636.18	37	FLT	1-month LIBOR	Monthly	1	1.657	1.397	1.397
Modified tenure	4,168.16	276,577.79	38	FLT	1-month LIBOR	Monthly	1	1.344	1.084	1.084
Modified term	38,071.44	515,419.12	38	FLT	1-month LIBOR	Monthly	1	1.384	1.124	1.124
Tenure	2,319.67	73,798.97	37	FLT	1-month LIBOR	Monthly	1	1.752	1.492	1.492
Term	30,407.98	144,900.42	37	FLT	1-month LIBOR	Monthly	1	1.260	1.000	1.000
Line of credit	4,128,213.10	88,895,122.30	38	FLT	1-month LIBOR	Monthly	1	0.998	0.738	0.738

Approximate weighted average gross lifetime interest rate cap(9)	Approximate weighted average MIP Fee(10)	approximate weighted average servicing fee margin(11)	Monthly servicing fee(12)	Monthly scheduled draw(13)	Approximate weighted average remaining draw term (in months)(14)	Available line of credit(15)	Maximum claim amount(16)	Pool number	HECM MBS issue date
13.275%	0.500%	0.060%	\$29,630.00	(18)	(18)	\$ 42,160,357.74	\$289,116,200.00	892359	December 2009
13.275	0.500	0.060	1,130.00	\$ 35,684.25	(19)	2,374,137.91	15,726,500.00	892359	December 2009
13.271	0.500	0.060	2,010.00	96,748.57	14	3,380,700.28	24,488,500.00	892359	December 2009
13.270	0.500	0.060	910.00	38,604.14	(19)	0.00	12,763,500.00	892359	December 2009
13.355	0.500	0.060	535.00	31,655.63	22	0.00	6,171,000.00	892359	December 2009
13.366	0.500	0.090	6,605.00	(18)	(18)	3,874,061.82	44,567,491.00	892371	December 2009
13.447	0.500	0.090	310.00	8,842.86	(19)	186,804.38	2,482,515.00	892371	December 2009
13.330	0.500	0.090	(17)	14,044.09	83	54,838.49	3,295,775.00	892371	December 2009
13.294	0.500	0.090	125.00	3,056.04	(19)	0.00	977,000.00	892371	December 2009
13.462	0.500	0.090	(17)	2,529.25	172	0.00	618,000.00	892371	December 2009
13.942	0.500	0.090	9,950.00	(18)	(18)	3,655,008.89	58,646,485.00	892373	December 2009
14.477	0.500	0.090	230.00	4,966.16	(19)	168,926.99	1,825,160.00	892373	December 2009
13.799	0.500	0.090	295.00	6,413.99	51	248,476.24	1,753,290.00	892373	December 2009
13.183	0.500	0.090	130.00	2,326.21	(19)	0.00	880,000.00	892373	December 2009
13.698	0.500	0.090	60.00	2,235.44	80	0.00	549,950.00	892373	December 2009
14.532	0.500	0.090	11,571.00	(18)	(18)	8,816,539.75	77,009,022.00	892374	December 2009
14.380	0.500	0.090	585.00	14,762.00	(19)	608,278.20	5,153,538.00	892374	December 2009
14.518	0.500	0.090	670.00	30,620.73	47	94,061.24	5,152,860.00	892374	December 2009
14.600	0.500	0.090	280.00	6,187.21	(19)	0.00	1,782,010.00	892374	December 2009
14.680	0.500	0.090	85.00	1,865.53	154	0.00	542,350.00	892373	December 2009
14.334	0.500	0.160	3,695.00	(18)	(18)	1,384,595.54	26,492,633.00	891840	June 2009
14.343	0.500	0.160	70.00	750.00	(19)	59,777.34	560,290.00	891840	June 2009
14.567	0.500	0.160	105.00	4,632.02	38	158,360.96	906,790.00	891840	June 2009
14.603	0.500	0.160	35.00	177.36	(19)	0.00	125,000.00	891840	June 2009
13.775	0.500	0.160	35.00	3,216.61	64	0.00	362,790.00	891840	June 2009
14.247	0.500	0.100	19,955.00	(18)	(18)	10,731,394.22	135,126,549.00	892600	November 2009

(continued)

Assumed characteristics of the HECMs and the participations underlying the trust assets(1)										
Payment plan	HECM MBS principal balance(2)	HECM loan balance	Approximate weighted average HECM age (in months) (3)	HECM interest type	Index	Rate reset frequency (4)	Next rate reset months (5)	Approximate weighted average gross interest rate (6)	Approximate weighted average gross margin(7)	approximate weighted average gross lifetime interest rate floor(8)
Modified tenure	185,396.54	2,521,431.96	39	FLT	1-month LIBOR	Monthly	1	0.958	0.698	0.698
Modified term	555,414.76	4,787,178.77	39	FLT	1-month LIBOR	Monthly	1	0.967	0.707	0.707
Tenure	228,230.16	3,377,186.82	38	FLT	1-month LIBOR	Monthly	1	0.986	0.726	0.726
Term	60,724.98	756,181.60	38	FLT	1-month LIBOR	Monthly	1	0.960	0.700	0.700
Line of credit	7,112,529.64	40,252,789.51	19	FLT	1-month LIBOR	Monthly	1	2.510	2.250	2.250
Modified tenure	335,160.64	1,660,646.47	18	FLT	1-month LIBOR	Monthly	1	2.510	2.250	2.250
Modified term	644,092.55	1,833,051.52	19	FLT	1-month LIBOR	Monthly	1	2.509	2.250	2.250
Tenure	799,042.10	1,666,493.11	19	FLT	1-month LIBOR	Monthly	1	2.510	2.250	2.250
Term	334,486.52	2,101,663.46	18	FLT	1-month LIBOR	Monthly	1	2.510	2.250	2.250
Line of credit	1,092,376.97	89,638,699.11	38	FLT	1-month LIBOR	Monthly	1	0.997	0.737	0.737
Modified tenure	52,182.02	2,567,623.77	38	FLT	1-month LIBOR	Monthly	1	0.974	0.714	0.714
Modified term	135,669.10	4,549,210.02	39	FLT	1-month LIBOR	Monthly	1	0.959	0.699	0.699
Tenure	59,286.97	3,513,477.70	38	FLT	1-month LIBOR	Monthly	1	0.977	0.717	0.717

Approximate weighted average gross lifetime interest rate cap(9)	Approximate weighted average MIP Fee(10)	approximate weighted average servicing fee margin(11)	Monthly servicing fee(12)	Monthly scheduled draw(13)	Approximate weighted average remaining draw term (in months)(14)	Available line of credit(15)	Maximum claim amount(16)	Pool number	HECM MBS issue date
14.629	0.500	0.100	775.00	14,283.14	(19)	786,666.47	6,253,210.00	892600	November 2009
14.631	0.500	0.100	1,320.00	31,017.08	79	259,105.80	9,311,740.00	892600	November 2009
14.560	0.500	0.100	995.00	21,332.02	(19)	0.00	7,959,515.00	892600	November 2009
14.370	0.500	0.100	230.00	5,434.84	98	0.00	1,514,790.00	892600	November 2009
12.498	0.500	0.060	8,575.00	(18)	(18)	17,037,176.20	85,095,300.00	892356	December 2009
12.500	0.500	0.060	420.00	14,829.97	(19)	722,937.75	6,017,000.00	892356	December 2009
12.497	0.500	0.060	375.00	40,475.05	14	729,977.43	5,290,500.00	892356	December 2009
12.497	0.500	0.060	420.00	13,379.10	(19)	0.00	5,171,000.00	892356	December 2009
12.498	0.500	0.060	310.00	23,134.31	18	0.00	4,546,000.00	892356	December 2009
14.142	0.500	0.100	20,250.00	(18)	(18)	11,057,553.19	136,880,385.00	892602	January 2010
14.285	0.500	0.100	810.00	15,150.88	(19)	813,466.84	6,503,210.00	892602	January 2010
14.539	0.500	0.100	1,285.00	31,017.08	89	257,897.61	8,948,950.00	892602	January 2010
14.498	0.500	0.100	1,055.00	22,341.88	(19)	0.00	8,304,675.00	892602	January 2010

Appendix 2: HECM PPC Curve: CPR Percentage in Effect by HECM Age

Age	CPR %	Age	CPR %	Age	CPR %	Age	CPR %	Age	CPR %	Age	CPR %
1	0.00000	61	15.09115	121	22.43490	181	29.77865	241	37.05000	301	40.05000
2	0.54545	62	15.21354	122	22.55729	182	29.90104	242	37.10000	302	40.10000
3	1.09091	63	15.33594	123	22.67969	183	30.02344	243	37.15000	303	40.15000
4	1.63636	64	15.45833	124	22.80208	184	30.14583	244	37.20000	304	40.20000
5	2.18182	65	15.58073	125	22.92448	185	30.26823	245	37.25000	305	40.25000
6	2.72727	66	15.70313	126	23.04688	186	30.39063	246	37.30000	306	40.30000
7	3.27273	67	15.82552	127	23.16927	187	30.51302	247	37.35000	307	40.35000
8	3.81818	68	15.94792	128	23.29167	188	30.63542	248	37.40000	308	40.40000
9	4.36364	69	16.07031	129	23.41406	189	30.75781	249	37.45000	309	40.45000
10	4.90909	70	16.19271	130	23.53646	190	30.88021	250	37.50000	310	40.50000
11	5.45455	71	16.31510	131	23.65885	191	31.00260	251	37.55000	311	40.55000
12	6.00000	72	16.43750	132	23.78125	192	31.12500	252	37.60000	312	40.60000
13	6.29167	73	16.55990	133	23.90365	193	31.24740	253	37.65000	313	40.65000
14	6.58333	74	16.68229	134	24.02604	194	31.36979	254	37.70000	314	40.70000
15	6.87500	75	16.80469	135	24.14844	195	31.49219	255	37.75000	315	40.75000
16	7.16667	76	16.92708	136	24.27083	196	31.61458	256	37.80000	316	40.80000
17	7.45833	77	17.04948	137	24.39323	197	31.73698	257	37.85000	317	40.85000
18	7.75000	78	17.17188	138	24.51563	198	31.85938	258	37.90000	318	40.90000
19	8.04167	79	17.29427	139	24.63802	199	31.98177	259	37.95000	319	40.95000
20	8.33333	80	17.41667	140	24.76042	200	32.10417	260	38.00000	320	41.00000
21	8.62500	81	17.53906	141	24.88281	201	32.22656	261	38.05000	321	41.05000
22	8.91667	82	17.66146	142	25.00521	202	32.34896	262	38.10000	322	41.10000
23	9.20833	83	17.78385	143	25.12760	203	32.47135	263	38.15000	323	41.15000
24	9.50000	84	17.90625	144	25.25000	204	32.59375	264	38.20000	324	41.20000
25	9.66667	85	18.02865	145	25.37240	205	32.71615	265	38.25000	325	41.25000
26	9.83333	86	18.15104	146	25.49479	206	32.83854	266	38.30000	326	41.30000
27	10.00000	87	18.27344	147	25.61719	207	32.96094	267	38.35000	327	41.35000
28	10.16667	88	18.39583	148	25.73958	208	33.08333	268	38.40000	328	41.40000
29	10.33333	89	18.51823	149	25.86198	209	33.20573	269	38.45000	329	41.45000
30	10.50000	90	18.64063	150	25.98438	210	33.32813	270	38.50000	330	41.50000
31	10.66667	91	18.76302	151	26.10677	211	33.45052	271	38.55000	331	41.55000
32	10.83333	92	18.88542	152	26.22917	212	33.57292	272	38.60000	332	41.60000
33	11.00000	93	19.00781	153	26.35156	213	33.69531	273	38.65000	333	41.65000
34	11.16667	94	19.13021	154	26.47396	214	33.81771	274	38.70000	334	41.70000
35	11.33333	95	19.25260	155	26.59635	215	33.94010	275	38.75000	335	41.75000
36	11.50000	96	19.37500	156	26.71875	216	34.06250	276	38.80000	336	41.80000
37	11.66667	97	19.49740	157	26.84115	217	34.18490	277	38.85000	337	41.85000
38	11.83333	98	19.61979	158	26.96354	218	34.30729	278	38.90000	338	41.90000
39	12.00000	99	19.74219	159	27.08594	219	34.42969	279	38.95000	339	41.95000
40	12.16667	100	19.86458	160	27.20833	220	34.55208	280	39.00000	340	42.00000

(continued)

Age	CPR %	Age	CPR %	Age	CPR %	Age	CPR %	Age	CPR %	Age	CPR %
41	12.33333	101	19.98698	161	27.33073	221	34.67448	281	39.05000	341	42.05000
42	12.50000	102	20.10938	162	27.45313	222	34.79688	282	39.10000	342	42.10000
43	12.66667	103	20.23177	163	27.57552	223	34.91927	283	39.15000	343	42.15000
44	12.83333	104	20.35417	164	27.69792	224	35.04167	284	39.20000	344	42.20000
45	13.00000	105	20.47656	165	27.82031	225	35.16406	285	39.25000	345	42.25000
46	13.16667	106	20.59896	166	27.94271	226	35.28646	286	39.30000	346	42.30000
47	13.33333	107	20.72135	167	28.06510	227	35.40885	287	39.35000	347	42.35000
48	13.50000	108	20.84375	168	28.18750	228	35.53125	288	39.40000	348	42.40000
49	13.62240	109	20.96615	169	28.30990	229	35.65365	289	39.45000	349	42.45000
50	13.74479	110	21.08854	170	28.43229	230	35.77604	290	39.50000	350	42.50000
51	13.86719	111	21.21094	171	28.55469	231	35.89844	291	39.55000	351	42.55000
52	13.98958	112	21.33333	172	28.67708	232	36.02083	292	39.60000	352	42.60000
53	14.11198	113	21.45573	173	28.79948	233	36.14323	293	39.65000	353	42.65000
54	14.23438	114	21.57813	174	28.92188	234	36.26563	294	39.70000	354	42.70000
55	14.35677	115	21.70052	175	29.04427	235	36.38802	295	39.75000	355	42.75000
56	14.47917	116	21.82292	176	29.16667	236	36.51042	296	39.80000	356	42.80000
57	14.60156	117	21.94531	177	29.28906	237	36.63281	297	39.85000	357	42.85000
58	14.72396	118	22.06771	178	29.41146	238	36.75521	298	39.90000	358	42.90000
59	14.84635	119	22.19010	179	29.53385	239	36.87760	299	39.95000	359	42.95000
60	14.96875	120	22.31250	180	29.65625	240	37.00000	300	40.00000	360	43.00000

Appendix 3: Example of Cashflow Payment Rules

(t starts from 1)

$$Rate_t(FA) = \text{Min}(Cap, LIBOR_t + \text{Margin})$$

$$Accrue Interest_t(FA) = Balance_{t-1}(FA) \times Rate_t(FA) / 12$$

$$(If \text{ Bond Principal}_t \geq 0)$$

$$Interest_t(FA) = \frac{Bond Interest_t \times Accrue Interest_t(FA)}{Bond Interest_t}$$

$$Principal_t(FA) = \frac{Bond Principal_t \times Balance_{t-1}(FA)}{Bond Balance_{t-1} - Bond Balance_0} + Balance_0(FA)$$

$$Interest_t(IA) = Bond Interest_t - Interest_t(FA)$$

$$(If \text{ Bond Principal}_t < 0)$$

$$Interest_t(FA) = \frac{(Bond Interest_t + Bond Principal_t) \times Accrue Interest_t(FA)}{Bond Interest_t}$$

$$Principal_t(FA) = 0$$

$$Interest_t(IA) = Bond\ Principal_t + Bond\ Interest_t - Interest_t(FA)$$

$$Principal_t(FA) = \min(Principal_t(FA), Balance_{t-1}(FA))$$

$$Balance_t(FA) = Balance_{t-1}(FA) + Accrue\ Interest_t(FA) - Interest_t(FA) - Principal_t(FA)$$

$$Balance_t(IA) = Bond\ Principal_t$$

Appendix 4: A Model for Estimating Repayment Speeds for HECMs

It is possible to estimate the repayment speeds on reverse mortgages by combining the mortality rates and mobility rates provided by the US Life Tables (see Rai 2009). The probability that a borrower will not die in the next 12 months, at any given age, can be represented by $(1 - \text{Mortality Rate [Annualized]})$. Similarly, the probability that a borrower does not move out in the next 12 months is equal to $(1 - \text{Mobility Rate [Annualized]})$.

Thus, the probability that the loan does not repay over the next 12 months is equal to:

$$(1 - \text{Mortality Rate}) * (1 - \text{Mobility Rate})$$

Therefore, the repayment speeds on reverse mortgages can be calculated as:

$$CPR = 1 - [(1 - \text{Mortality Rate [Annualized]}) * (1 - \text{Mobility Rate [Annualized]})]$$

Estimating repayments based on just mortality and mobility rates provides reasonable results, though the model will tend to overestimate repayment speeds as compared to the repayment speeds based on historical data, due to self-selection. Borrowers who are in good health are more likely to take out reverse mortgage loans despite high origination costs.

References

- Alexander, L. *The root cause of the current liquidity/credit crunch*. Citigroup, December 4, 2008. Department of Housing and Urban Development.
- Ginnie Mae REMIC Offering Circulars 2010-0100, June 23, 2010.
- Ginnie Mae MBS Guide. Chapter 35. Home equity conversion mortgage loan pools – special requirements (HECM//HMBS), July 1, 2011.
- HUD, Handbook 4235.1 REV-1, November 18, 1994.
- NRMLA – Consumer site administered by the National Reverse Mortgage Lenders Association. URL: <http://www.nrmlaonline.org/rms/>

- Patterson, M., V. Barberio, G. Bailey, and M. Kelsch. *Rating U.S. home equity conversion mortgage securitizations*. Fitch ratings, June 4, 2007.
- Rai, V., T. Garg, B. Rose, and R. Parulekar. *A primer on reverse mortgages*. Citigroup, August 11, 2009.
- Rao, J., B. Gordon, and Z. Koscielniak. *A primer on reverse mortgages*. Morgan Stanley, October 6, 2009.
- Schultz, G. *Understanding home equity reverse mortgages*. Wells Fargo Securities, July 12, 2010.
- Schultz, G., and R. Ahlgren. *Understanding the H-REMIC waterfall*. Wells Fargo Securities, July 21, 2010.
- Strand, N., and S. Deb. *Reverse mortgages and HMBS*. Barclays Capital, March 18, 2010.



An Analysis of Risk Treatment in the Field of Finance

60

Fernando Gómez-Bezares and Fernando R. Gómez-Bezares

Abstract

This chapter studies the way to introduce risk into financial analysis. It suggests and analyzes metrics to improve the classic methods, like the Sharpe and Treynor ratios or the Net Present Value.

In particular, the chapter suggests a linear penalization method developed by the authors, applying it to performance indexes and capital budgeting, and considering both total and systematic risk. The method shows obvious advantages in comparison with traditional methods.

Keywords

Capital budgeting · Linear penalization · Performance · Risk

JEL classification

G11 · G31

This chapter is an updated version of the work published by the same authors in the Second Edition of the *Encyclopedia of Finance* by Springer, pp. 705–711.

F. Gómez-Bezares
Deusto Business School, Bilbao, Spain
e-mail: f.gomez-bezares@deusto.es

F. R. Gómez-Bezares (✉)
Mutua Madrileña, Madrid, Spain
e-mail: fgomezbezares@mutua.es

60.1 Introduction

It has been generally accepted¹ in the field of finance that individuals are risk averse. This implies that risk must be considered in order to value an asset or to measure the performance of an investment. In short words, those assets that are riskier need to have a higher return (on average) to be considered an attractive investment. Both the way in which risk is measured and how to introduce it as an input into analyses have been key subjects of financial research, this has been particularly true since the middle of the twentieth century.

Indeed, the risk problem has been widely studied in financial literature, where we can mention the State-Preference Theory² as a general framework. However, in the fields of stock market performance and capital budgeting techniques, the most applied models until now are those derived from the Portfolio Theory,³ like the Capital Asset Pricing Model (CAPM)⁴ of Sharpe (1964), Lintner (1965), and Black (1972).

This chapter brings up a clear problem that appears on performance indexes as commonly used as Sharpe's or Treynor's, questioning the classifications they make. We think that the root of the problem is to use the quotient to penalize for risk, suggesting the linear penalization as an alternative. This gives birth to a new system of performance measuring, the PIRR, which in one of its versions is equivalent to Jensen's α index.

Classic performance indexes, like Sharpe's, Treynor's, and Jensen's, have been and currently are widely applied both by practitioners (in their asset management) and by academics (in their research). Just to quote a few, authors like Chua and Koh (2007) apply Sharpe; Hodges et al. (2003) apply Treynor; Sainz et al. (2006) or Fama and French (2010) apply Jensen; Yu and Liow (2009), Benefield et al. (2009), Lipton and Kish (2010), or Mazumder et al. (2010) apply several of these indexes in varied contexts. Also can be mentioned: Collins and Fabozzi (2000), In et al. (2008), Chu and McKenzie (2008), Lee et al. (2009), or Ho et al. (2011). These indexes are being applied recently and very often, for example, Zulkafli et al. (2017) or Anam et al. (2020). This shows the relevance of present paper, as it develops advantageous alternatives to Sharpe's and Treynor's indexes, suggesting as well an improvement in Jensen's index.

In our opinion, indexes must be easy to calculate, and above all easy to understand. This is why we suggest very practical indexes that besides are measured in returns (this is one of the advantages of Modigliani and Modigliani's index despite it matches its classification with Sharpe's index). On the other hand, our proposed

¹Although some exceptions have also been pointed out.

²It can be seen in Copeland et al. (2005), based on works like Arrow (1964) and Debreu (1959).

³Markowitz (1952, 1959).

⁴Although asset valuation models based on real options are also being increasingly applied for the last years, we are not entering here into their discussion.

indexes avoid the problems before mentioned. Of course that linear penalization also has some problems, but we consider them less important, as we will explain later.

The linear penalization system can also be extended on to NPV with a clearly original development, giving rise to the PPV,⁵ which is commented within the chapter.

In this work we are not going to question the hypothesis of whether individuals are risk averse, neither the way to measure this risk⁶ (total or systematic risk within the CAPM framework). What we are going to discuss is the way in which risk must be introduced into analyses. With this goal, we will start by reviewing the investment classification implied by the Sharpe and Treynor ratios, and we will suggest an alternative method. We will also analyze how to penalize the Net Present Value, ending with some conclusions.

60.2 Sharpe's Classification

When measuring the performance of a portfolio, and presuming its liquidity, we must consider return and risk. Within the context of the Portfolio Theory, we will understand risk as variability, and we will measure it with the standard deviation. This is more justifiable assuming a normal distribution of returns, as the normal distribution is defined with the mean⁷ (μ) and standard deviation (σ) parameters. This approach is consistent with Sharpe (1966), who suggests classifying portfolios according to their performance with the following ratio:

$$S = \frac{\mu - r_0}{\sigma} \quad (1)$$

where μ and σ are the mean and standard deviation of portfolio returns, and r_0 is the risk-free rate. Assuming a map on $\mu - \sigma$ axes, where μ_m and σ_m are the mean and standard deviation for the Market Portfolio returns (R_m), and within the framework of the Portfolio Theory, we find the Capital Market Line (CML) of Fig. 1. We can also identify portfolios R_a and R_b with their respective means and deviations. It is easy to confirm that the tangents of the α angles in this figure are the respective Sharpe ratios of the three portfolios.⁸

⁵ The NPV is considered, in a classic reference in finance, as is Brealey et al. (2008), as the first one of the seven most important ideas in finance. In order to introduce risk into the NPV, the most widely applied systems are the *risk-adjusted rate of return valuation formula* and the *certainty equivalent valuation formula* (Copeland et al. 2005, pp. 156–157). We will use an alternative system, derived from the NPV: the *Penalized Present Value (PPV)*.

⁶ Although both subjects have been and will be topic of discussion.

⁷ Expected value.

⁸ And we will explain later the meaning of the PIRR.

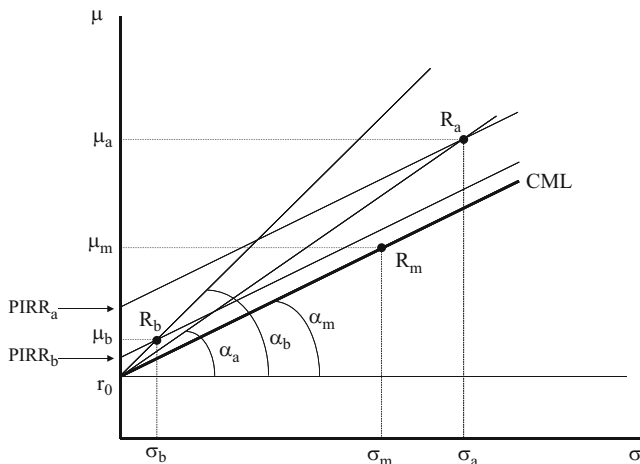


Fig. 1 Shows the Capital Market Line and values of the Sharpe ratio and the PIRR

The Sharpe ratio is undoubtedly one of the most employed indexes in the financial industry, and its portfolio classification coincides with the ranking that we can derive from the M^2 suggested by Modigliani and Modigliani (1997). However, it has a significant flaw: presuming normal distributions of returns, Sharpe classifies portfolios on the basis of their probability of giving a return below r_0 , which is easy to verify simply by looking at formula (1). We consider that in many cases this ranking cannot be supported: a portfolio π may have a much higher μ than another portfolio ϕ , but if ϕ had a slightly smaller probability of giving a return below r_0 , it would be classified with Sharpe or M^2 better than π , despite the fact that nearly always π will give much higher returns than ϕ .

The following example illustrates this flaw of the Sharpe ratio:

We can assume that we have two investment funds, “A” and “B,” with the following mean (μ) and standard deviation (σ) for their returns:

$$\mu_A = 30\% \quad \sigma_A = 9\%$$

$$\mu_B = 7\% \quad \sigma_B = 1\%$$

If the risk-free rate is $r_0 = 3\%$, the Sharpe ratios for these investment funds are:

$$S_A = (30 - 3)/9 = 3 \quad S_B = (7 - 3)/1 = 4$$

Following the Sharpe ratio, “B” investment fund is better, as its S is higher. If we assume that returns follow a normal distribution, the probability of “B” giving a return below the risk-free rate (3%) is 0.00003. The same probability for “A” is 0.00135. As Sharpe ranks investments minimizing this probability, then $S_B > S_A$.

However, it seems quite clear that common investors would prefer “A” investment fund. In fact, **with a probability of ~99.5%, the returns of “A” investment**

fund will be higher than the average return of “B” investment fund. Therefore, in this type of cases, Sharpe’s ratio would be misleading to take investment decisions.⁹

60.3 Treynor’s Classification

Starting from the well-known differentiation between systematic and diversifiable risk:

$$\sigma_j^2 = \beta_j^2 \sigma_m^2 + \sigma_\varepsilon^2 \quad (2)$$

Where we can read: the total risk of portfolio j (measured here as the variance of its returns) is equal to the square of its beta multiplied by the market variance, plus the variance of ε (that represents diversifiable risk). Assuming now that the portfolio whose performance we are measuring will be integrated into a wider and well diversified portfolio, the variance of ε can be disregarded (as diversification will make it null), and we can measure the risk with beta.¹⁰ This leads us to the Treynor (1965) ratio¹¹:

$$T = \frac{\mu - r_0}{\beta} \quad (3)$$

where beta is the measure of systematic risk in the CAPM. In Fig. 2, quite similar to number 1, we find the Security Market Line (SML), in a CAPM μ - β map, where we can see how the Treynor ratio of each portfolio is also measured as the tangent of its respective α angle.¹²

If we now replace formula (2)¹³ in formula (1), and disregarding the variance of ε ¹⁴:

$$\begin{aligned} S &= \frac{\mu - r_0}{\sqrt{\beta^2 \sigma_m^2 + \sigma_\varepsilon^2}} \approx \frac{\mu - r_0}{\sqrt{\beta^2 \sigma_m^2}} = \frac{\mu - r_0}{\beta \sigma_m} \\ &= \frac{T}{\sigma_m} \end{aligned} \quad (4)$$

⁹It might be argued that somebody could, by borrowing money at the risk-free rate, start from “B” and reach an investment that outperformed “A,” and this is true. However, small investors cannot get themselves into debt at the risk-free rate. Besides, the performance of an investment fund is known ex-post, and by then, it is impossible for any investor to leverage his or her investment.

¹⁰As the market variance will be equal for all the portfolios.

¹¹The classification obtained is the same as with M^2 for beta (Modigliani 1997).

¹²We will explain the PIRR for beta later.

¹³Dispensing here with the j subindex.

¹⁴For this reason we are employing the $\approx$ symbol.

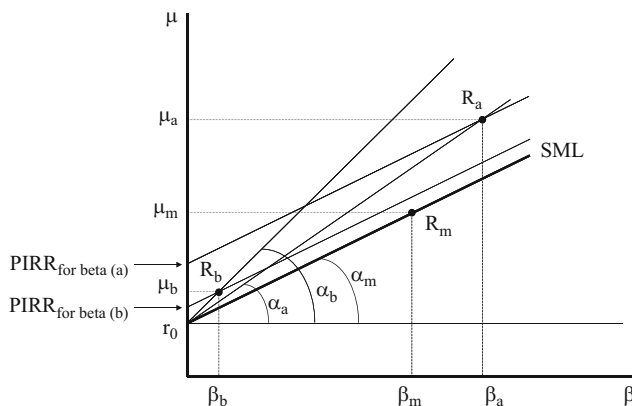


Fig. 2 Shows the Security Market Line and values of the Treynor ratio and the PIRR for beta

Considering that σ_m is equal for every portfolio, Sharpe and Treynor would lead us to the same ranking, and this is what happens for well-diversified portfolios.

We can go further and assume that a non-diversified portfolio whose performance we are measuring is going to be integrated into a wider and well diversified portfolio, and that consequently the diversifiable risk will disappear.¹⁵ In this case, the probability of giving a return below r_0 (which is measured by Sharpe as we mentioned above) will only depend on beta.¹⁶ **We can thus say that Treynor also ranks portfolios according to their probability of giving a return below r_0 , and we have already criticized this.**

60.4 Advantages of Linear Penalization: PIRR and PIRR for Beta

The paradoxes identified in Sharpe's and Treynor's indexes come from applying a system of risk quotient penalization. Gómez-Bezares (1993) suggested a linear penalization system that was applied to the case we are facing by Gómez-Bezares et al. (2004). The Penalized Internal Rate of Return (PIRR) is calculated by deducting t standard deviations (σ) from the expected return (μ) of the analyzed portfolio:

$$PIRR = \mu - t\sigma \quad (5)$$

The PIRR value is a performance metric that should be interpreted as a certainty equivalent for a portfolio with risk, being therefore useful to compare the risk-adjusted performance of different portfolios. The return premium asked for each risk unit would be t , and can be understood as given by the market: t would be the market Sharpe (the tangent of α_m in Fig. 1). In fact, replacing the parameters of the market portfolio in Eq. 5

¹⁵This would allow to measure risk with β and the use of Treynor.

¹⁶Out of the components of total risk, the diversifiable risk will disappear, and the market standard deviation is the same for every portfolio. Therefore, the only difference will be β .

we get to $PIRR_m = r_0$. In consequence, r_0 would be the certainty equivalent of the market portfolio, this should be interpreted by saying that individuals, on average, are indifferent between r_0 and R_m , which seems very reasonable.¹⁷

On the other hand, and assuming normal distributions, t has a clear statistical meaning: for a t value, the PIRR would be the guaranteed minimum return with a determined probability (that would depend on t value). This system is equivalent to the use of parallel indifference straight lines, as can be confirmed in Fig. 1, where can also be seen that in this case the portfolio classification derived from the use of the PIRR differs from that one derived from Sharpe.¹⁸ Therefore, the PIRR consists of a different system to Sharpe's index, despite both methods consider the total risk of investments.

We can see now what happens when applying the PIRR to the same example we developed in Sect. 60.2. Assuming $t = 1$ we will have (values in %):

$$PIRR_A = 30 - 9 = 21$$

$$PIRR_B = 7 - 1 = 6$$

Being "A" investment fund clearly better, which is the same conclusion that mere common sense suggested. From a statistical point of view, assuming a normal distribution of returns, a value of $t = 1$ leaves a probability of 0.15866; therefore, in this case, the calculated PIRR is a guaranteed minimum return with a probability of 0.84134. As mentioned before, it is very logical to take the market Sharpe as t value, but investors could be occasionally more demanding.

Supposing now that the diversifiable risk can be disregarded, and using as t the Sharpe of the market, we will get to the following expression:

¹⁷It could also be reasoned in this way: the market Sharpe, which is the slope of the CML, shows the "price of risk," which is the increase in the expected return that is demanded for a risk increase. Consequently, it is logical to give such a value to t in the formula (5).

¹⁸The use of indifference straight lines is a constraint of the system, which can be mitigated by the use of increasing t values as the risk grows. Sharpe also employs indifference straight lines, but instead of being parallels, they form a beam.

In Fig. 1, following the Sharpe ratio, all the portfolios located on the same straight line starting from r_0 would be indifferent. For example, if we focus on the straight line linking r_0 and R_a , and according to the Sharpe ratio, all the portfolios on this straight line are indifferent to R_a , because their Sharpe value is the same: the tangent of α_a . At the same time, these portfolios would be worse than those located on the straight line linking r_0 and R_b , because the latter ones have a higher Sharpe. This is the consequence of Sharpe using a system of indifference straight lines forming a beam.

In the case of the PIRR system, indifference straight lines are drawn in parallel to the CML. The crossing point with the vertical axis is the PIRR, and its value is bigger the higher the straight line crosses with the vertical axis.

If we applied increasing t values, as it was mentioned before, we would get closer to obtain indifference curves and we could reach to a solution quite similar to the standard one.

$$\begin{aligned}
 PIRR &= \mu - t\sigma = \mu - \frac{\mu_m - r_0}{\sigma_m} \sigma \\
 &= \mu - \frac{\mu_m - r_0}{\sigma_m} \sqrt{\beta^2 \sigma_m^2 + \sigma_\varepsilon^2} \\
 &\approx \mu - (\mu_m - r_0)\beta
 \end{aligned} \tag{6}$$

The PIRR obtained in Eq. 6 can be named **PIRR for beta**, and this would be the performance metric alternative to Treynor. Its functioning can be seen in Fig. 2, and much of what has been said when comparing the PIRR with Sharpe can here be repeated and adapted. PIRR for beta leads us to the same portfolio classification as the Jensen α (1968, 1969).

Therefore, both PIRR and PIRR for beta have the following advantages in comparison with Sharpe's and Treynor's indexes, respectively:

1. Their values can be **easily interpreted** as a certainty equivalent of average return, with the degree of confidence that investors want to demand in order to take their investment decisions. These return values measured by PIRR and PIRR for beta are much easier to understand by an average investor.
2. These metrics based on linear penalization **avoid the risk shown in the example** of Sect. 60.2, and do not rank investments by their probability of giving a return below r_0 .

Therefore, there are clear differences between the PIRR and the Sharpe ratio, and also between the PIRR for beta and the Treynor ratio. In the case of Jensen α , it classifies investments in the same way as the PIRR for beta does, but the PIRR for beta value is once again easier to interpret. The PIRR method is being applied, as it can be seen, for example, in Ferruz et al. (2009, 2010), Irimia-Diéguez et al. (2014), or Santibáñez et al. (2020).

60.5 Penalized Present Value

The Penalized Present Value (PPV) suggested by Gómez-Bezares (1993) is based on Net Present Value ¹⁹ (NPV), and penalizes μ_{NPV} (expected NPV) according to σ_{NPV} in order to use this result as a decision criterion.

$$PPV = \mu_{NPV} - t \sigma_{NPV} \tag{7}$$

This is a certainty equivalent, on which similar considerations can be made to those already set out for the PIRR.²⁰

¹⁹Discounted at the risk-free rate, as the risk penalization is applied afterwards.

²⁰Use of indifference straight lines, statistical meaning of t . . . Regarding t value, it should grow with the investment period as well as the market Sharpe tends to do so, as can be seen in Sharpe (1994).

Assuming now a unique period of time, and investing an amount I into project i (with expected return μ_i and risk σ_i), we will get to:

$$\mu_{NPV}(i) = \frac{I(1 + \mu_i)}{1 + r_0} - I = I \frac{\mu_i - r_0}{1 + r_0} \quad (8)$$

The standard deviation of NPV will be σ (of return), multiplied by the investment and discounted at the risk-free rate. It is easy to see what the result is when the investment return is r_0 , in this case μ_{NPV} would be zero. We can transform Fig. 1, making a scale change in the vertical axis, so that appears μ_{NPV} and we get to Fig. 3. The PPV_i will be calculated deducting from μ_{NPV} the slope that now has the CML multiplied by σ_i . We see in Eq. 9 that these operations result in μ_{NPV} minus the market Sharpe multiplied by σ_{NPV} .²¹

$$\begin{aligned} PPV(i) &= I \frac{\mu_i - r_0}{1 + r_0} - \frac{I \frac{\mu_m - r_0}{1 + r_0}}{\sigma_m} \sigma_i \\ &= I \frac{\mu_i - r_0}{1 + r_0} - \frac{\mu_m - r_0}{\sigma_m} \sigma_{NPV} \end{aligned} \quad (9)$$

We can also see PPV_i as $\sigma_{NPV} \{ \sigma_{NPV} = I \cdot \sigma_i / (1 + r_0) \}$ multiplied by the difference of Sharpes: first half of Eq. 10. In the second half of Eq. 10 we find (between square brackets) the difference between the obtained premium and the expected premium

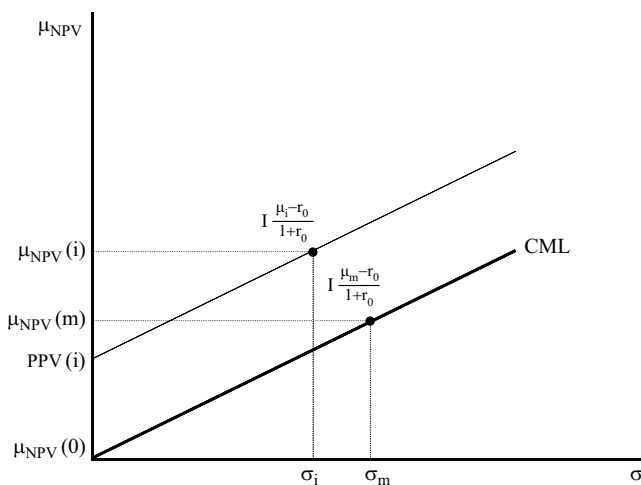


Fig. 3 Shows the Penalized Present Value (PPV)

²¹ Here, as in some other occasions, we dispense with i subindex for NPV, not to make notation more complicated.

according to the CML of Fig. 1. Multiplying this extra premium by I , and discounting at the risk-free rate, we also get to PPV_i .

$$\begin{aligned} PPV(i) &= \sigma_{NPV} \left(\frac{\mu_i - r_0}{\sigma_i} - \frac{\mu_m - r_0}{\sigma_m} \right) \\ &= \frac{I}{1 + r_0} \left[(\mu_i - r_0) - \frac{\mu_m - r_0}{\sigma_m} \sigma_i \right] \end{aligned} \quad (10)$$

Making in the vertical axis of Fig. 2 a scale change like the one applied between Figs. 1 and 3, we will get to Fig. 4, where we find the PPV for beta. Starting from the first half of Eq. 9, if we disregard the diversifiable risk, we will have:

$$\begin{aligned} PPV \text{ for beta}(i) &= I \frac{\mu_i - r_0}{1 + r_0} - I \frac{\mu_m - r_0}{1 + r_0} \beta_i \\ &= \frac{I}{1 + r_0} [(\mu_i - r_0) - (\mu_m - r_0) \beta_i] \end{aligned} \quad (11)$$

This can also be read in Fig. 4 as μ_{NPV} minus the slope of SML multiplied by beta. In the second half of Eq. 11 we find the extra return premium obtained,²² which multiplied by I and discounted at the risk-free rate also gives the PPV for beta.

The β_{NPV} will be equal to β_i multiplied by I and discounted at the risk-free rate:

$$\beta_{NPV}(i) = \frac{\sigma_{NPV,m}}{\sigma_m^2} = \frac{\sigma_{i,m}}{\sigma_m^2} * \frac{I}{1 + r_0} = \beta_i \frac{I}{1 + r_0} \quad (12)$$

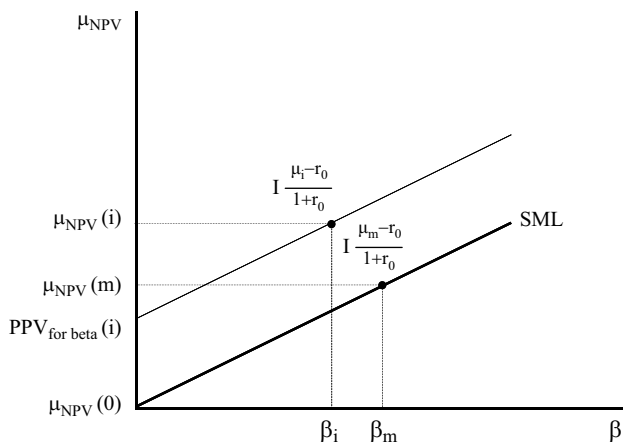


Fig. 4 Shows the Penalized Present Value (PPV) for beta

²²This is the Jensen α .

where $\sigma_{i,m}$ is the covariance between i return and market return. It can be deduced from the first half of Eq. 11 that:

$$\begin{aligned} PPV \text{ for } \beta(i) &= I \frac{\mu_i - r_0}{1 + r_0} - I \frac{\mu_m - r_0}{1 + r_0} \beta_i \\ &= \mu_{NPV}(i) - (\mu_m - r_0) \beta_{NPV}(i) \end{aligned} \quad (13)$$

This coincides with the classic certainty equivalent valuation formula.²³

60.6 Final Thoughts and Conclusions

This chapter presents two models of linear penalization of returns that may be of use (one model with σ and another one with β), as well as the penalization of NPV. It also shows the consistency between models in a way that we consider very original.

The system of linear penalization is an interesting alternative to Sharpe's or Treynor's indexes, as its output **can be easily interpreted as a certainty equivalent, and avoids certain mistakes in performance evaluation that Sharpe or Treynor ratios could cause.**

However, it is known that in this type of models the most correct option is to use indifference curves. In this sense, our model is a simplification, as it uses straight lines. This criticism may also be applied to the Sharpe and Treynor ratios, as they use a beam of indifference straight lines and consider indifferent all those portfolios placed on the same straight line. Such criticism can be mitigated in our PIRR model by applying increasing “ r ” values, and therefore getting closer to use indifference curves. Working on this, we could face other potential criticisms related to the use of parallel indifference straight lines, because it is not the same to beat the CML (or the SML) in a certain amount for a low risk than for a high risk, being easier in the latter this extraordinary performance to be by chance. We consider that practicality and simplicity of use of the developed models should be important targets to consider, and this is why we think that proposed simplifications would be appropriate in many occasions.

Applying this methodology to asset valuation, the PPV may be an interesting criterion for capital budgeting with risk. In this chapter is explained in more detail its application for a single investment period, but it can also be generalized to n periods. An elementary way to do so could be assuming reinvestments at the risk-free rate, therefore converting a case of n periods into a single period case.²⁴

Within this chapter, it has been assumed that penalization “ t ” value was given by the market. However, this “ t ” value could be modified by the analyst, and then all the proposed methods would have new possibilities. Applying this option to the PPV would be particularly interesting. In order to extend the possibilities to apply PPV

²³It can be adapted, for instance, from Copeland et al. (2005, p. 157).

²⁴It would also be possible to reinvest at the WACC (Weighted Average Cost of Capital) rate, despite this would take us to apply a larger t value than in the previous case.

and PIRR, we refer the reader to Gómez-Bezares (1993) and to Gómez-Bezares and Gómez-Bezares (2015), respectively.

Acknowledgments We want to thank the editor, referees, and other colleagues for their comments and suggestions, which have undoubtedly contributed to improve the present entry.

References

- Anam, K., S. Hadi, N.F. Kusumawati and E.P. Suwanto. 2020. Evaluate the performance of the stock portfolio by oil palm companies using Sharpe, Treynor and Jensen index. *IOSR Journal of Economics and Finance*, series III, July-August, 7–14. Open Access: <https://www.iosrjournals.org/iosr-jef/pages/vol.11i4%20Series-3.html>
- Arrow, K.J. 1964. The role of securities in the optimal allocation of risk-bearing. *Review of Economic Studies* 31 (2): 91–96.
- Benefield, J.D., R.I. Anderson, and L.V. Zuppano. 2009. Performance differences in property-type diversified versus specialized real estate investment trusts (REITs). *Review of Financial Economics* 18: 70–79.
- Black, F. 1972. Capital market equilibrium with restricted borrowing. *Journal of Business* 45 (3): 444–455.
- Brealey, R.A., S.C. Myers, and F. Allen. 2008. *Principles of corporate finance*. 9th ed. New York: McGraw-Hill.
- Chu, P.K., and M. McKenzie. 2008. A study on stock-selection and market-timing performance: evidence from Hong Kong Mandatory Provident Funds (MPF). *Review of Pacific Basin Financial Markets and Policies* 4: 617–649.
- Chua, C.T., and W.T.H. Koh. 2007. Measuring investment skills of fund managers. *Applied Financial Economics* 17: 1359–1368.
- Collins, B., and F. Fabozzi. 2000. Equity manager selection and performance. *Review of Quantitative Finance and Accounting* 15: 81–97.
- Copeland, T.E., J.F. Weston, and K. Shastri. 2005. *Financial theory and corporate policy*. 4th ed. Boston: Pearson Addison Wesley.
- Debreu, G. 1959. *The theory of value*. New York: Wiley.
- Fama, E.F., and K.R. French. 2010. Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance* 65: 1915–1947.
- Ferruz, L., F. Gómez-Bezares, and M. Vargas. 2009. Performance measures: advantages of linear risk penalization. *Applied Financial Economics* 19: 73–85.
- . 2010. Portfolio theory, CAPM and performance measures. In *Handbook of quantitative finance and risk management*, ed. C.-F. Lee, A.C. Lee, and J. Lee, 267–281. New York: Springer.
- Gómez-Bezares, F. 1993. Penalized present value: net present value penalization with normal and beta distributions. In *Capital budgeting under uncertainty*, ed. R. Aggarwal, 91–102. Englewood Cliffs: Prentice-Hall.
- Gómez-Bezares, F., and F.R. Gómez-Bezares. 2015. Don't use quotients to calculate performance. *Cogent Economics and Finance* 3 (1065584): 1–14. Open Access: <http://www.tandfonline.com/doi/full/10.1080/23322039.2015.1065584>.
- Gómez-Bezares, F., J.A. Madariaga, and J. Santibáñez. 2004. Performance ajustada al riesgo: índices clásicos y nuevas medidas. *Análisis Financiero* 93: 6–16.
- Ho, L.-C., J. Cadle, and M. Theobald. 2011. An analysis of risk-based asset allocation and portfolio insurance strategies. *Review of Quantitative Finance and Accounting* 36: 247–267.
- Hodges, C.W., W.R.L. Taylor, and J.A. Yoder. 2003. Beta, the Treynor ratio, and long-run investment horizons. *Applied Financial Economics* 13: 503–508.

- In, F., S. Kim, V. Marisetty, and R. Faff. 2008. Analysing the performance of managed funds using the wavelet multiscaling method. *Review of Quantitative Finance and Accounting* 31: 55–70.
- Irimia-Diéguez, A.I., J. Bernal, and M.D. Oliver-Alfonso. 2014. The financial performance of an innovative megaproject. *Procedia – Social and Behavioral Sciences* 119: 417–426. Open Access: <https://www.sciencedirect.com/journal/procedia-social-and-behavioral-sciences/vol/119/suppl/C>.
- Jensen, M.C. 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23 (2): 389–416.
- . 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* 42 (2): 167–247.
- Lee, C.-F., K. Wang, and Y.L. Chen. 2009. Hedging and optimal hedge ratios for international index futures markets. *Review of Pacific Basin Financial Markets and Policies* 4: 593–610.
- Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47 (1): 13–37.
- Lipton, A.F., and R.J. Kish. 2010. Robust performance measures for high yield bond funds. *Quarterly Review of Economics and Finance* 50: 332–340.
- Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7 (1): 77–91.
- . 1959. *Portfolio selection: efficient diversification of investments*. New York: Wiley.
- Mazumder, M.I., E.M. Miller, and O.A. Varela. 2010. Market timing the trading of international mutual funds: weekend, weekday and serial correlation strategies. *Journal of Business Finance and Accounting* 37: 979–1007.
- Modigliani, L. 1997. *Don't pick your managers by their alphas*. U.S. Investment Research, U.S. Strategy, Morgan Stanley Dean Witter.
- Modigliani, F., and L. Modigliani. 1997. Risk-adjusted performance. *Journal of Portfolio Management, Winter* 23 (2): 45–54.
- Sainz, J., P. Grau, and L.M. Doncel. 2006. Mutual fund performance and benchmark choice: the Spanish case. *Applied Financial Economics Letters* 2: 317–321.
- Santibáñez, J., L. Goitia-Berriozabal, and J.A. Madariaga. 2020. Estructura financiera y creación de valor: una aproximación a través de la TRIP. *Boletín de Estudios Económicos* 75: 143–170.
- Sharpe, W.F. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* 19 (3): 425–442.
- . 1966. Mutual fund performance. *Journal of Business* 39 (1): 119–138.
- . 1994. The Sharpe ratio. *Journal of Portfolio Management, Fall* 21 (1): 49–58.
- Treynor, J.L. 1965. How to rate management of investment funds. *Harvard Business Review* 43: 63–75.
- Yu, S.M., and K.H. Liow. 2009. Do retail firms benefit from real estate ownership? *Journal of Property Research* 26: 25–60.
- Zulkafli, A.H., Z. Ahmad, and M. Eky Ermal. 2017. The performance of socially responsible investments in Indonesia: a study of the Sri Kehati Index (SKI). *Gadjah Mada International Journal of Business* 19 (1): 59–76. Open Access: <https://jurnal.ugm.ac.id/gamaijb/article/view/17959/15682>.



The Trading Performance of Dynamic Hedging Models: Time Varying Covariance and Volatility Transmission Effects

61

Michael T. Chng and Gerard L. Gannon

Abstract

In this chapter, we investigate the value of incorporating implied volatility from related option markets in dynamic hedging. We comprehensively model the volatility of all four S&P 500 cash, futures, index option and futures option markets simultaneously. Synchronous half-hourly observations are sampled from transaction data. Special classes of extended simultaneous volatility systems (ESVL) are estimated and used to generate out-of-sample hedge ratios. In a hypothetical dynamic hedging scheme, ESVL-based hedge ratios, which incorporate incremental information in the implied volatilities of the two S&P 500 option markets, generate profits from interim rebalancing of the futures hedging position that are incremental over competing hedge ratios. In addition, ESVL-based hedge ratios are the only hedge ratios that manage to generate sufficient profit during the hedging period to cover losses incurred by the physical portfolio.

Keywords

Volatility transmission · Dynamic hedging · Optimal hedge ratio · S&P 500

JEL Classification

G14 · G28

61.1 Introduction

The size of a hedge is affected by the sensitivity of the return of the underlying asset to the return of the hedging instrument. This sensitivity measure β is also called the optimal hedge ratio (OHR). The hedge ratio is deemed optimal in terms of minimizing the

M. T. Chng (✉) · G. L. Gannon

School of Accounting, Economics and Finance, Deakin University, Melbourne, VIC, Australia

e-mail: mchng@deakin.edu.au

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,

https://doi.org/10.1007/978-3-030-91231-4_61

1411

variability in the value of the overall position. A dynamic hedging scheme recognizes that since the OHR varies over time i.e. $\beta = \beta_t$, the hedging outcome can be improved from interim rebalancing of the hedging position. An issue addressed by the hedging literature is the incremental performances of competing OHRs over a static hedge. If the corresponding variance and covariance terms used to calculate the OHR can be adequately modeled and forecasted, then hedging performance should improve.

The volatility literature contains a voluminous debate on the information content of implied volatility versus historical volatility. The general consensus is that some combination of both improves volatility forecasts. Our main objective in this chapter is to formally blend the two literatures together. We investigate if incremental information from combining implied and historical volatilities translates into incremental hedging performance by the corresponding OHR. We consider the use of the S&P 500 index futures (henceforth FI) contract to hedge against a widely-held portfolio that tracks the S&P 500 index e.g. S&P Depository Receipts (SPDR). In this chapter, we assume the S&P 500 cash index (henceforth CI) as our physical portfolio.¹ Denote N_t as the optimal number of futures contracts to short against an existing long position in the underlying portfolio at time t . This is calculated in Eq. 1.

$$N_t = \beta_t \times \frac{V_{CI,t}}{V_{FI,t}} \quad (1)$$

$$\beta_t = \frac{\sigma_{CI,FI,t}}{\sigma_{FI,t}^2} = \rho_{CI,FI,t} \times \frac{\sigma_{CI,t}}{\sigma_{FI,t}}$$

$V_{CI,t}$ and $V_{FI,t}$ represent the value of the physical portfolio to be hedged against and the value covered by each futures contract respectively. This size-ratio $\frac{V_{CI,t}}{V_{FI,t}}$ is adjusted by β_t , which is calculated as the covariance $\sigma_{CI,FI,t}$ between the spot return $r_{CI,t}$ and futures returns $r_{FI,t}$ divided by the variance of the futures return $\sigma_{FI,t}^2$. As β_t varies during the hedging period, so does N_t , which implies a need to rebalance the hedging position. Since $\rho_{CI,FI,t} = \frac{\sigma_{CI,FI,t}}{\sigma_{CI,t} \times \sigma_{FI,t}}$, N_t can be expressed as $\rho_{CI,FI,t} \times \frac{\sigma_{CI,t}}{\sigma_{FI,t}}$, where $\sigma_{CI,t}$ is the cash index volatility. If we assume both $V_{CI,t}$ and $V_{FI,t}$ are exogenous, the OHR is the only parameter to estimate to determine the optimal N_t .² This demonstrates the importance of modeling $\sigma_{CI,FI,t}$ and $\sigma_{FI,t}^2$ in a dynamic hedging scheme. If we assume constant perfect positive correlation i.e. $\rho_{CI,FI,t} = 1$, modeling and forecasting the OHR is analogous to modeling and forecasting $\sigma_{CI,t}$ and $\sigma_{FI,t}$.

The preceding highlights an intimate link between the literatures on hedging and volatility modeling. Nonetheless, potential volatility transmission between spot, futures and options markets are often ignored in the OHR estimation. This is despite an incumbent literature spanning over 15 years debating the merits of implied versus historical volatility. If incremental information embedded in implied volatility exists,

¹S&P Depository Receipts are yet to exist during our sample period.

²Strictly speaking, $V_{CI,t}$ is not entirely exogenous in that a decision is required on the proportion of the underlying asset's value to hedge against. However, this is a separate issue from the OHR determination.

then incorporating such volatility transmissions should generate better volatility forecasts and improve the OHR estimation. The S&P 500 CI and FI each possesses a well-established option market. The S&P 500 futures option (FO) trading pit is located beside the futures pits on the Chicago Mercantile Exchange (CME) trading floor. The S&P 500 index option (IO) is traded on the Chicago Board Options Exchange (CBOE).

Pennings and Meulenberg (1997) provide a comprehensive review of hedging performance measures in the literature. The evaluation generally contrasts between the combined cash-futures position versus the cash position alone. Hedging performance is measured by the reduction in the variance of the combined position in Ederington (1979), the ratio of Sharpe ratios (cash-futures position divided by cash position only) in Howard and D'Antonio (1984), difference in certainty equivalent returns between the combined versus cash position in Hsin et al. (1994) and expected utility maximization in Kroner and Sultan (1993). Hedging applications based on a bivariate GARCH framework include Baillie and Myers (1991) for US commodity futures, Kroner and Sultan (1993) for foreign currency futures and Park and Switzer (1995) for US stock index futures. In these studies, hedging performance is based on minimizing the variance of the overall position.

Lee et al. (2000) and Yeh and Gannon (2000) evaluate out-of-sample volatility estimates generated from competing volatility models in the context of a dynamic hedging scheme. However, potential volatility transmission effects were not investigated, but transaction costs were included in the analysis in the latter paper. Chng and Gannon (2003) model contemporaneous volatility and volume effects within a framework of formalized structure of simultaneous volatility equations proposed by Gannon (1994). The order and matrix rank conditions for the simultaneous volatility and competing misspecified volatility models are documented in this latter paper. While the resultant volatility forecasts statistically dominate those from competing volatility models, an out-of-sample evaluation based on market inference was not performed.

Au-Yeung and Gannon (2005) employ a M-GARCH model modified to allow for multiple structural breaks and volatility spillovers between the HSIF cash and index futures markets and also overnight spillovers from the S&P 500 index futures. Gannon (2005) repeated this analysis but employed a Full Information Maximum Likelihood (FIML) set of simultaneous volatility equations. Bhattacharya et al. (2007) employed the Au-Yeung and Gannon M-GARCH model to test volatility spillovers between Indian stock and share futures. Au-Yeung and Gannon (2008) repeated the early analysis of structural breaks in their M-GARCH model and showed that inclusion of volume of trade effects led to insignificant structural break parameters. Lee et al. (2009) utilize four static and a dynamic bivariate GARCH models to find the OHR for the S&P 500 and five major Asian market index futures. Lee et al. (2010) investigate hedge ratios in international futures markets in the light of the cross-country linkage and interaction using a 3SLS estimation procedure. Gannon (2010) re-visited the simultaneous volatility class of models to test transmission and spillover effects when the intra-day sampling interval reduces. In this chapter further theoretical conditions for this class of models

and systems Error Correction terms are defined. In all of the above estimators, hedge ratios can be extracted but no out of sample hedging performance was undertaken.

We evaluate the hedging performances of competing OHRs base on the incremental profits from interim rebalancing that each OHR generates over a static hedge. We argue that our performance measure is more consistent with the investigation on whether incorporating incremental information from implied volatility in the OHR calculation translates into incremental profits from rebalancing the futures position. We evaluate eight competing OHRs generated from competing volatility and covariance forecasts.

Our secondary objective is to model the entire time-varying variance-covariance matrix in a formal system of simultaneous volatility (SVL) equations. This can be seen as a competing estimator to variants of the restricted class of bivariate GARCH estimators. However, we consider only the constant correlation bivariate GARCH and univariate GARCH estimators as similar applications in the published literature have report improved performance over regression-based estimators. The true unrestricted estimator of the simultaneous multivariate GARCH (MGARCH) hedge ratio is generated from the VEC-MGARCH form. This cannot be readily applied to model time varying covariance due to convergence problems.³ As noted in the literature for the BEKK-GARCH and dynamic constant correlation forms of Engle (2002) and Tse and Tsui (2002), the covariance term from these models are functions of conditional standard deviations of equations in the system.

The BEKK-GARCH estimator proposed by Engle and Kroner (1995) overcomes the convergence problem by employing an estimator that guarantees a positive definite variance/covariance matrix. The base BEKK estimator is still restricted as it employs a function of the product of the respective bivariate time varying standard deviations as estimators of the time varying co-variance term. Various extensions of this class of estimator allow for functions of the standard deviations to generate the covariance term. Some in sample estimators also allow for asymmetric positive and negative return effects (sign effect) in the estimation process. One drawback with employing the asymmetric versions in out of sample forecasting applications is lack of future values of the “sign effect” to generate the forecasts.

In this chapter, we utilize market data from parallel option markets to capture market anomalies, including sign effects, because these are continuously observable variables rather than discrete imposed indicator variables. This can also be seen as a benchmark for the full structural volatility system estimator that incorporates market

³The unrestricted vector GARCH specification seldom leads to identifiable point estimates since doing so requires the inversion of the variance-covariance matrix at each sample point. When off-diagonal covariance terms are large relative to diagonal variance terms, the determinant of the matrix tends to zero, such that the inverse matrix can be unidentified. Such cases are very likely in the context of spot and the futures returns. This technical problem is commonly addressed by imposing a constant correlation and focus on modeling time varying spot and futures volatility, which somewhat defeats the purpose of dynamic hedging.

transmission effects. It could also be seen as a benchmark for the aforementioned class of BEKK-GARCH estimators within the comparison framework of this chapter and an area for future research. Clearly the real focus in this chapter is to compare two alternative versions of the class of simultaneous volatility estimators SVL and ESVL. Comparisons with other estimators is restricted to those reported using similar evaluation processes.

The rest of this chapter proceeds as follow. Institutional features and sampling procedures are provided in Sect. 61.2. In Sect. 61.3, competing volatility models and various methods of computing the OHR are discussed. Model estimation results are reported in Sect. 61.4. Section 61.5 reports the out-of-sample hedging performance. Section 61.6 concludes.

61.2 Institutional Background and Data Sampling

The analysis in our paper requires a synchronously sampled set of intraday data for all four contracts. We found it difficult to obtain a set from both options markets that was satisfactory. We decided to employ the same cash index and index futures dataset for the results reported by Miller et al. (1994). For this dataset, a full database of both options transactions is available.⁴

The S&P 500 CI is a value-weighted broad-based market index that comprises 500 widely-held stocks.⁵ It is regarded by the financial community as a barometer to gauge the performance of the US equity market. The S&P 500 FI follows a March, June, September and December contract cycle. In 1990, the contract multiplier was \$500 per index point. The tick size is 0.05 point, or \$25. On average, the contracts are rolled over between the 6th and 11th of the contract month. For our sample, we choose the 8th day of the delivery month as the date to switch to the next contract.

The S&P 500 FO are American options. One futures option is written on one futures contract and is quoted in index points. The tick size is 0.05.⁶ These contracts follow a monthly cycle, and data is available across a range of contract months and strike prices.⁷ The two nearest-to-maturity option contracts (e.g. Feb and March options written on the March futures contract) will trade at strike prices in multiples

⁴The data was supplied by TICKDATA Inc. but for various reasons, they stopped collection of options data streams.

⁵Individual stock prices are multiplied by the number of shares outstanding. The products are summed up and standardized by a pre-determined base value. Base values for the index are adjusted to reflect changes in capitalization due to mergers, acquisitions and rights issues etc.

⁶However, a trade may occur at a price of 0.025 index point if it is necessary to liquidate positions to allow for both parties to trade.

⁷For the March futures contract, there are option contracts expiring at the end of Jan, Feb, March etc. up to May. Options expiring prior or during the March quarterly cycle are written on the March futures contracts. Else, they are written on the June futures contract.

of 5. Longer maturity contracts will trade at strike prices in multiples of 10. The option is exercised at maturity if it is in the money. Both the S&P 500 FI and FO contracts are traded under the Index and Option Market Division on the CME. For each trading day, the average futures price of the front contract is used to determine the closest to money FO contract to include in our sample. The implied volatility from the FO will be the average of the implied volatility between the call and put. The S&P 500 IO contracts share similar contract specifications to the FO contracts. A key difference is that the IO are cash-settled and they are European options.

From 4th Jan to 31st December 1990, we sample near-synchronous 30-min observations from the nearest-to-maturity contracts for each of the four markets. For all the markets, normal trading commences at 8.30 a.m. and finishes at 3.15 p.m. To avoid mechanically-induced opening and closing effects, we use the 8.40 a.m. and 3 p.m. prices to compute the 9 a.m. opening return and 3 p.m. closing return correspondingly.⁸ When extracting implied volatilities, potential non-synchronous spot and option prices violate the specification of the Black-Scholes pricing model, which dilutes the validity of subsequent implied volatility measures. However, since both option markets are highly liquid, non-synchronicity between spot and option prices is not a major problem for our study. We calculate an option's term-to-maturity as $\frac{\text{No of trading days to maturity}}{\text{Trading days per annum}}$.⁹ The continuously compounded 1-month Treasury bill rate is used to proxy the risk-free rate.¹⁰ Lastly, the continuous-compounding annualized dividend yield daily time-series of the CI is used to back out implied volatility of the IO.

61.3 Volatility Estimators and Competing Hedge Ratios

Denote $r_{CI,t}$ and $r_{FI,t}$ as the half-hourly continuously compounded returns of CI and FI respectively. We construct CI and FI volatilities as absolute returns $\sigma_{CI,t} = |r_{CI,t}|$ and $\sigma_{FI,t} = |r_{FI,t}|$. The covariance between CI and FI is defined as $\sigma_{CI,FI,t} = |r_{CI,t}| |r_{FI,t}|$.¹¹

⁸For missing observations in one market, corresponding observations from other markets are excluded. This gives a total of 3186 out of a possible 3289 observations. There are 13 half-hourly observations per day over 253 trading days in 1990. Details of the sampling procedure can be obtained from the authors upon request.

⁹For example, an option trading on 23rd April with May as the delivery month will have 27 trading days till maturity. Following conventions, this study uses 252 days for the denominator. Thus for that option, term-to-maturity is calculated as $23/252 = 0.0913$.

¹⁰This is chosen over the 30 year Treasury bond rate, which may contain term structure premium.

¹¹Although negative covariance is possible and should be allowed for, the spot and futures returns are expected to move in the same direction most of the time due to cost-of-carry. An examination of the data confirms this statement.

61.3.1 Constructing Out-of-Sample Hedge Ratios

To generate out-of-sample OHR forecasts, we divide our 1-year sample into two halves. The first half is our estimation sample and contains 1613 observations. The second half, which is our test sample, contains 1573 observations. Out-of-sample OHR projections into the test period are derived by sequentially expanding the estimation period one observation at a time to update the coefficient estimates and generate a series of one-step ahead OHR forecasts. We evaluate the hedging performance of eight hedge ratios that are constructed from different volatility sources and models. These are summarized in Table 1.

The first hedge ratio is obtained from a least square regression of $r_{CI,t}$ against $r_{FI,t}$ and a dummy variable Dum^{Close} for market closing effects.¹² Outlined in Eq. 2, this β_{OLS} measure is commonly used in a static hedge, where no re-balancing occurs during the hedging period. We denote the coefficient β_{OLS} as OHR(1).

$$r_{CI,t} = \beta_0 + \beta_1 (Dum^{Close}) + \beta_{OLS}(r_{FI,t}) + \varepsilon_t \quad (2)$$

OHR(2) is constructed from the conditional volatility of univariate GARCH(1,1) estimations for $\sigma_{CI,t}$ and $\sigma_{FI,t}$. In Eq. 3, $\varepsilon_{CI,t}$ and $\varepsilon_{FI,t}$ are residuals obtained from the CI and FI return-equations. Lastly, $\rho_{CI,FI}$ denotes the correlation between $r_{CI,t}$ and $r_{FI,t}$.

$$\begin{aligned} \sigma_{CI,t+1}^2 &= \alpha_{10} + \alpha_{11}(\varepsilon_{CI,t}^2) + \beta_{11}(\sigma_{CI,t}^2) \\ \sigma_{FI,t+1}^2 &= \alpha_{20} + \alpha_{21}(\varepsilon_{FI,t}^2) + \beta_{21}(\sigma_{FI,t}^2) \\ OHR(2) &= \rho_{CI,FI} \times \frac{\sigma_{CI,t+1}}{\sigma_{FI,t+1}} \end{aligned} \quad (3)$$

OHR(3) is designed to be an enhanced version of OHR(2). Here, we examine the potential incremental information provided by an array of additional variables added to the CI and FI GARCH (1,1) variance equations. These include $\sigma_{FI,t}^2$ in the $\sigma_{CI,t+1}^2$

Table 1 The list of competing hedge ratios

OHR (1)	Naïve hedge
OHR (2)	Constant covariance bivariate GARCH(1,1)
OHR (3)	Constant covariance bivariate (enhanced) GARCH(1,1)
OHR (4)	Constant covariance IHR
OHR (5)	Constant covariance ESVL(CI) and ESVL(FI)
OHR (6)	SVL
OHR (7)	Time varying covariance IHR
OHR (8)	Time varying covariance SVL, ESVL(CI) and ESVL(FI)

¹² $Dum^{Close} = 1$ for the 3.00 p.m. closing return, and 0 otherwise.

equation, $\sigma_{CI,t}^2$ in the $\sigma_{FI,t+1}^2$ equation, $\sigma_{IO,t}^2$ and $\sigma_{FO,t}^2$. In addition, we also allow potential volume effects to enter the conditional variance equations. Denote CV_{t-1} and FV_{t-1} as the corresponding change in CI and FI tick-volume.¹³ Lastly, we include Dum^{Open} ¹⁴ and Dum^{Close} for market opening and closing effects. If there exists any incremental information in the implied volatility and/or tick volumes that can be adequately brought out through a GARCH framework, OHR(3) should outperform OHR(2).

$$\begin{aligned}
 \sigma_{CI,t+1}^2 &= \alpha_{10} + \alpha_{11} \left(\varepsilon_{CI,t}^2 \right) + \beta_{11} \left(\sigma_{CI,t}^2 \right) + \gamma_{11} (Dum^{Open}) \\
 &\quad + \gamma_{12} (Dum^{Close}) + \gamma_{13} \sigma_{FI,t}^2 + \gamma_{14} \sigma_{IO,t}^2 \\
 &\quad + \gamma_{15} \sigma_{FO,t}^2 + \gamma_{16} (CV_t) \\
 \sigma_{FI,t+1}^2 &= \alpha_{20} + \alpha_{21} \left(\varepsilon_{FI,t}^2 \right) + \beta_{21} \left(\sigma_{FI,t}^2 \right) + \gamma_{21} (Dum^{Open}) \\
 &\quad + \gamma_{22} (Dum^{Close}) + \gamma_{23} \sigma_{CI,t}^2 + \gamma_{24} \sigma_{IO,t}^2 \\
 &\quad + \gamma_{25} \sigma_{FO,t}^2 + \gamma_{26} (FV_t)
 \end{aligned}$$

$$OHR(3) = \rho_{CI,FI} \times \frac{\sigma_{CI,t+1}}{\sigma_{FI,t+1}} \quad (4)$$

For the actual estimation, we begin with a comprehensive GARCH(1,1) specification outlined in Eq. 4. This includes all volatility and tick-volume variables from related markets. First, we exclude non-estimable variables from the weighting series of the variance equations. Next, we use both the Wald test statistics and Akaike Information Criterion (AIC) to guide us as to which variables to exclude in an effort to systematically exclude variables from the comprehensive model. The trimmed down GARCH(1,1) specification is then used to generate out-of-sample projections for OHR(3).

Implied volatility is regarded by option traders as a forward-looking estimate. Accordingly, a hedge ratio can be constructed simply with $\sigma_{IO,t}$ and $\sigma_{FO,t}$. We term this as the implied hedge ratio (IHR), which is a novel alternative to compute hedge ratios by substituting $\sigma_{CI,t}$ with $\sigma_{IO,t}$ and $\sigma_{FI,t}$ with $\sigma_{FO,t}$. Since both $\sigma_{IO,t}$ and $\sigma_{FO,t}$ vary over time, IHR is applicable to a dynamic hedging scheme. In Eq. 5, we label IHR as OHR(4).

$$OHR(4) = IHR = \rho_{CI,FI} \times \left(\frac{\sigma_{IO,t}}{\sigma_{FO,t}} \right) \quad (5)$$

¹³Tick volume is defined as the number of half hourly price changes. For example, CV_t is the number of half-hourly price changes between time $t-1$ and t .

¹⁴ $Dum^{Open} = 1$ for the 9.00 a.m. opening return, and 0 otherwise.

Since we consider only nearest to money put and call options of the front contract, and since both option markets are liquid, non-synchronicity between spot and option prices should not be a major concern. The Black-Scholes model adjusted for continuously compounded dividend yield is used to compute $\sigma_{IO,t}$ from the observed premium of the European index options, assuming dividends are non-stochastic. However, Black's (1976) model for pricing futures options cannot be readily applied to back out $\sigma_{FO,t}$ as these are American options. If the premium for early exercise is non-trivial, then this should be acknowledged.¹⁵ However, Whaley (1986) finds that early exercise premium exists only for in-the-money S&P 500 futures put options. Since $\sigma_{FO,t}$ is computed as the average of nearest-to-maturity call and put options, we postulate that any potential upward biases in $\sigma_{FO,t}$ from not explicitly adjusting for early exercise premium should be trivial.

OHR(5) is generated from a four-equation extended simultaneous volatility (ESVL) framework that incorporates cross-market volatility transmissions across all four S&P 500 markets. The ESVL system is based on the three-equation SVL model proposed in Gannon (1994). Since OHR(5) requires forecasts for both $\sigma_{CI,t}$ and $\sigma_{FI,t}$, two separate ESVL systems are considered, where $\sigma_{CI,t}$ and $\sigma_{FI,t}$ is each specified as the variable to be forecasted. We refer to these as ESVL(CI) and ESVL(FI) respectively. In specifying the structure of both ESVL systems, we assume that the FO possesses the greatest informational efficiency, followed concurrently by IO and FI. The cash index ranks last in terms of efficiency ranking. This is confirmed from pairwise Granger-causality tests on the volatility time series. ESVL(CI) is presented in Eq. 6. Diagnostic tests indicate a significant opening effect for the cash index. As such, an opening dummy Dum^{Open} is included in the estimation of ESVL(CI).

$$\begin{aligned}
 \sigma_{CI,t} &= \alpha_{10} + \alpha_{11}Dum^{Open} + \beta_{11}(\sigma_{FI,t}) + \beta_{12}(\sigma_{IO,t}) \\
 &\quad + \gamma_{11}(\sigma_{CI,t-1}) + \gamma_{12}(\sigma_{FI,t-1}) + \gamma_{13}(CV_{t-1}) + \varepsilon_{CI,t} \\
 \sigma_{FI,t} &= \alpha_{20} + \alpha_{21}Dum^{Open} + \beta_{21}(\sigma_{CI,t}) + \beta_{22}(\sigma_{FO,t}) \\
 &\quad + \gamma_{21}(\sigma_{CI,t-1}) + \gamma_{22}(\sigma_{FI,t-1}) + \gamma_{23}(\sigma_{FO,t-1}) + \varepsilon_{FI,t} \\
 \sigma_{IO,t} &= \alpha_{30} + \alpha_{31}Dum^{Open} + \beta_{31}(\sigma_{CI,t}) + \beta_{32}(\sigma_{FO,t}) \\
 &\quad + \gamma_{31}(\sigma_{CI,t-1}) + \gamma_{32}(\sigma_{IO,t-1}) + \gamma_{33}(\sigma_{FO,t-1}) + \varepsilon_{IO,t} \\
 \sigma_{FO,t} &= \alpha_{40} + \alpha_{41}Dum^{Open} + \beta_{41}(\sigma_{CI,t}) + \beta_{42}(\sigma_{FI,t}) \\
 &\quad + \gamma_{41}(\sigma_{CI,t-1}) + \gamma_{42}(\sigma_{FI,t-1}) + \gamma_{43}(\sigma_{FO,t-1}) + \varepsilon_{FO,t}
 \end{aligned} \tag{6}$$

Since $\sigma_{CI,t}$ is the variable to be forecasted, it is present in all four equations of ESVL(CI) to ensure that a reduced-form can be obtained for generating out-of-sample forecasts. Lagged tick-volume CV_{t-1} is included in the $\sigma_{CI,t}$ equation to overcome a singularity estimation problem. The futures market leads the spot market due to lower execution costs, higher liquidity and more informed trading. As such,

¹⁵ While it can be shown that early exercise is never optimal for a futures option if the premium is subjected to futures-style margining, the premiums for FO are paid up front. With positive interest rates, early exercise remains a possibility.

$\sigma_{FI,t}$ is specified to enter the $\sigma_{CI,t}$ equation. Similarly, $\sigma_{FO,t}$ is specified to enter both the $\sigma_{FI,t}$ and $\sigma_{IO,t}$ equations. $\sigma_{CI,t-1}$ and $\sigma_{FI,t-1}$ are included in the $\sigma_{IO,t}$ and $\sigma_{FO,t}$ equations to account for possible volatility feedback from the underlying assets back to the corresponding option market.

The derivation of the reduced-form ESVL(CI) is provided in the appendix. The coefficients of the reduced-form $\sigma_{CI,t}$ equation are functions of the structural coefficients from ESVL(CI). These are estimated using the estimation sample. Together with time t volatility variables, we can generate a one-step ahead forecast $\sigma_{CI,t+1}$ according to the specification of the reduced-form $\sigma_{CI,t}$ equation.¹⁶ Conducting one-step ahead forecasts based on the reduced-form $\sigma_{CI,t}$ equation allows inherent volatility transmission effects inherent among all four S&P 500 markets to be incorporated into $\sigma_{CI,t+1}$. After each forecast, the coefficients are sequentially updated one observation at a time. This recursive process generates a time-series of $\sigma_{CI,t+1}$.¹⁷

ESVL(FI) is presented in Eq. 7. A closing dummy Dum^{Close} is included as the FI market displays a significant closing effect. The series of one-step ahead forecasts $\sigma_{FI,t+1}$ from ESVL(FI) are obtained in a similar fashion to $\sigma_{CI,t+1}$. OHR(5) is presented in Eq. 8.

$$\begin{aligned}
 \sigma_{FI,t} &= \alpha_{10} + \alpha_{11}Dum^{Close} + \beta_{11}(\sigma_{CI,t}) + \beta_{12}(\sigma_{IO,t}) \\
 &\quad + \gamma_{11}(\sigma_{FI,t-1}) + \gamma_{12}(\sigma_{FO,t-1}) + \gamma_{13}(FV_{t-1}) + \varepsilon_{CI,t} \\
 \sigma_{CI,t} &= \alpha_{20} + \alpha_{21}Dum^{Close} + \beta_{21}(\sigma_{FI,t}) + \beta_{22}(\sigma_{IO,t}) \\
 &\quad + \gamma_{21}(\sigma_{FI,t-1}) + \gamma_{22}(\sigma_{CI,t-1}) + \gamma_{23}(\sigma_{IO,t-1}) + \varepsilon_{FI,t} \\
 \sigma_{IO,t} &= \alpha_{30} + \alpha_{31}Dum^{Close} + \beta_{31}(\sigma_{FI,t}) + \beta_{32}(\sigma_{FO,t}) \\
 &\quad + \gamma_{31}(\sigma_{FI,t-1}) + \gamma_{32}(\sigma_{IO,t-1}) + \gamma_{33}(\sigma_{FO,t-1}) + \varepsilon_{IO,t} \\
 \sigma_{FO,t} &= \alpha_{40} + \alpha_{41}Dum^{Close} + \beta_{41}(\sigma_{FI,t}) + \beta_{42}(\sigma_{CI,t}) \\
 &\quad + \gamma_{41}(\sigma_{FI,t-1}) + \gamma_{42}(\sigma_{CI,t-1}) + \gamma_{43}(\sigma_{FO,t-1}) + \varepsilon_{FO,t}.
 \end{aligned} \tag{7}$$

¹⁶ An alternative way to view the mapping of reduced form parameters back to the structural form parameters is to think about the systems in terms of the normalized matrix rank condition rather than the unnormalized matrix rank condition. In the former, there is an implied unity restriction imposed on endogenous variables in own structural equations. Then substitution between the structural and reduced forms is straightforward and in an identifiable system all structural parameters in every structural equation are identified. As such, the reduced-form equations provide unique projections of endogenous variables in the systems. However, there can be cases where alternative structural parameterizations provide non-nested competing sets of structural systems. In this case, artificial nested testing procedures need be employed to select between competing systems. In our paper, there are competing identifiable three and four equation systems. These are compared in terms of out of sample hedge ratio estimation and subsequently in terms of trade re-balancing performance.

¹⁷ To note, utilizing ESVL-based hedge ratios may seem computationally tedious. However, as the forecasts are made one-step at a time, the majority of coefficients are stable when we sequentially expand the estimation period. Once the initial coefficient estimates are recorded, subsequent updating is computationally easy.

$$OHR(5) = \rho_{CI,FI} \times \left(\frac{\sigma_{ESVL(CI),t+1}}{\sigma_{ESLV(FI),t+1}} \right) \quad (8)$$

OHR(6) is based on the modeling of a time-varying variance-covariance matrix of the spot and futures markets based on the SVL system in Eq. 9. Two separate reduced-form equations with $\sigma_{FI,t}$ and $\sigma_{CI,FI,t}$ as the variable to be forecasted are derived from (9). These are used to generate a series of one-step ahead forecasts for $\sigma_{FI,t}$ and $\sigma_{CI,FI,t}$.¹⁸ The presence of $\sigma_{CI,FI,t}$ in the $\sigma_{FI,t}$ equation is driven by the fact that during non-volatile trading periods, the basis is expected to be small, reflecting cost-of-carry. But during volatile trading periods, the futures market is expected to experience a greater increase in trading activity relative to the spot market, causing a transitory deviation between the two. This implies both a decrease in $\sigma_{CI,FI,t}$ and an increase in $\sigma_{FI,t}$.

$$\begin{aligned} \sigma_{FI,t} &= \alpha_{10} + \alpha_{11}Dum^{Close} + \beta_{11}(\sigma_{CI,FI,t}) + \gamma_{11}(\sigma_{FI,t-1}) \\ &\quad + \gamma_{12}(\sigma_{CI,FI,t-1}) + \gamma_{13}(FV_{t-1}) + \varepsilon_{FI,t} \\ \sigma_{CI,t} &= \alpha_{20} + \alpha_{21}Dum^{Close} + \beta_{21}(\sigma_{FI,t}) + \gamma_{21}(\sigma_{FI,t-1}) \\ &\quad + \gamma_{22}(\sigma_{CI,t-1}) + \gamma_{23}(\sigma_{FI,CI,t-1}) + \varepsilon_{CI,t} \\ \sigma_{CI,FI,t} &= \alpha_{30} + \alpha_{31}Dum^{Close} + \beta_{31}(\sigma_{FI,t}) + \gamma_{31}(\sigma_{FI,t-1}) \\ &\quad + \gamma_{32}(\sigma_{CI,t-1}) + \gamma_{23}(\sigma_{FI,CI,t-1}) + \varepsilon_{FI,CI,t} \end{aligned} \quad (9)$$

$$OHR(6) = \frac{\sigma_{CI,FI,t+1}}{\sigma_{FI,t+1}^2} \quad (10)$$

As our preceding argument does not imply anything about the direction of causality, we allow $\sigma_{FI,t}$ to enter the $\sigma_{CI,FI,t}$ equation. OHR(6) is presented in Eq. 10. Compared to other hedge ratios, OHR(6) does not incorporate any volatility transmissions from related options markets. However, it does consider time-varying covariance.

$$OHR(7) = \frac{\sigma_{CI,FI,t+1}}{\sigma_{FO,t+1}^2} \quad (11)$$

$$OHR(8) = \frac{\sigma_{CI,FI,t+1}}{\sigma_{ESVL,FI,t+1}^2} \quad (12)$$

The last two hedge ratios are based on hybrids projections from ESVL and/or SVL. Outlined in Eq. 11, OHR(7) combines time-varying covariance forecasts generated from Eq. 9 with potential incremental information from the implied variance of the FO market. OHR(8) is presented in Eq. 12. Here, we combine $\sigma_{CI,FI,t}$ forecasts generated from the SVL with $\sigma_{FI,t}$ forecasts generated from ESVL(FI). As such, OHR(8) is the only hedge ratio in this chapter that incorporates both spot-futures time-varying covariance and volatility transmission from related option markets.

¹⁸The derivations are available upon request.

61.4 Empirical Results

61.4.1 Descriptive Statistics and Preliminary Results

In Table 2, descriptive statistic for $r_{CI,t}$ and $r_{FI,t}$ are presented in Panel 1, the autocorrelation of key variables in Panel 2, the correlation matrix of volatility and volume variables in Panel 3 and stationarity test statistics in Panel 4.

To note, $r_{FI,t}$ is highly significant at the 13th lag. Since our sample consists of 13 half-hourly return observations per trading day, the significant 13th lag could reflect time of the day effect that is evident in most index futures markets. The strong correlation between $\sigma_{CI,t}$ and $\sigma_{FI,t}$ is expected. Both $\sigma_{CI,t}$ and $\sigma_{FI,t}$ are also correlated with CV_t and FV_t respectively. In contrast, the correlations between $\sigma_{CI,t}$ and $\sigma_{IO,t}$ as well as between $\sigma_{FI,t}$ and $\sigma_{FO,t}$ are comparatively weaker. Lastly, $\sigma_{IO,t}$ and $\sigma_{FO,t}$ are not strongly correlated with each other, suggesting that the two option markets are not necessarily close substitutes. Augmented Dickey Fuller (ADF) test statistics indicate that cash index, futures prices and their tick volumes are integrated of order one I(1), such that $r_{CI,t}$, $r_{FI,t}$, CV_t and FV_t are all stationary. The volatility time series of the four S&P markets are also tested and found to be stationary.

61.4.2 Results from Model Estimates

First, we report GARCH estimation results. Potential higher-order GARCH-effects are investigated using nested tests, but additional parameters are found to be insignificant for both $\sigma_{CI,t}$ and $\sigma_{FI,t}$. The final GARCH specification for the CI and FI are determined from a ‘top-down’ approach based on Wald tests statistics and the AIC to systematically exclude variables from the weighting series in the GARCH variance equation. As such, the specification of $\sigma_{CI,t}$ is slightly different from $\sigma_{FI,t}$. Both are presented below with the coefficient estimates and significance levels.

$$\begin{aligned}\sigma_{CI,t}^2 &= 0.000^{a**} + 0.1065^{**} \left(\varepsilon_{CI,t-1}^2 \right) + 0.7989^{**} \left(\sigma_{CI,t-1}^2 \right) \\ &\quad + 0.000^{**} (Dum^{Open}) + 0.0001^{**} (CV_{t-1}) + 0.0001^{**} (FV_{t-1}) \\ \sigma_{FI,t}^2 &= 0.000 + 0.0389 \left(\varepsilon_{FI,t-1}^2 \right) + 0.8857^{**} \left(\sigma_{FI,t-1}^2 \right) \\ &\quad + 0.000^* (Dum^{Close}) + 0.021 \left(\sigma_{IO,t-1}^2 \right) + 0.0001 (CV_{t-1}) \\ &\quad + 0.0001^{**} (FV_{t-1})\end{aligned}$$

*a: * indicate significance at 5% level; ** indicate significance at 1% level.*

Next, we discuss results from ESVL and SVL estimations. We perform two diagnostic tests to improve the validity of the ESVL specifications. First, we test for significant autocorrelations in the residuals, which is indicative of omission of other relevant variables. Our preliminary analysis reveals significant auto-correlation

Table 2 Descriptive statistics of return, volatility and volume

Panel A: Descriptive statistics									
	Mean	Min	Max	Median	Std dev	Skewness	Excess kurtosis		
$r_{CL,t}$	0.0000058	− 0.02057	0.030119	0.000058	0.002156	− 0.29727	15.27724		
$r_{FL,t}$	0.0019101	− 0.037979	0.10251	0.000148	0.009579	5.31734	33.55719		
Panel B: Autocorrelation features									
Lag	PACF ^a $r_{CL,t}$	Q(k) of DX ^b	PACF $r_{FL,t}$	Q(k) of DX ²	PACF $\sigma_{CL,t}$	Q(k) of DX	Q(k) of DX	PACF $\sigma_{FL,t}$	PACF $\sigma_{FL,t}$
1	0.0978 ^c	60.2 ^c	− 0.0476	14.2 ^c	0.184 ^c	213 ^c	− 0.0132	1.09	
2	0.0171	64.6 ^c	− 0.0277	18.3 ^c	0.113 ^c	342 ^c	− 0.0153	2.54	
5	0.0197	67.2 ^c	− 0.0500 ^c	48.3 ^c	0.0277 ^c	508 ^c	− 0.0421 ^c	25.9 ^c	
10	0.0362 ^c	83.3 ^c	− 0.0589 ^c	96.7 ^c	0.0395 ^c	615 ^c	0.0415 ^c	74.9 ^c	
13	0.0117	85.8 ^c	0.929 ^c	5580 ^c	0.0644 ^c	827 ^c	0.959 ^c	5,890 ^c	
26	0.0086	105 ^c	0.437 ^c	11,100 ^c	0.1011 ^c	1,380 ^c	0.386 ^c	11,700 ^c	
Panel C: Correlation matrix									
$\sigma_{CL,t}^2$	$\sigma_{CL,t}^2$	$\sigma_{FL,t}^2$	$\sigma_{FO,t}^2$	$\sigma_{IO,t}^2$	CV_t	FV_t	dCV_t	dFV_t	
$\sigma_{CL,t}^2$	1								
$\sigma_{FL,t}^2$	0.2280	1							
$\sigma_{FO,t}^2$	− 0.0233	− 0.0054	1						
$\sigma_{IO,t}^2$	0.0722	− 0.0009	− 0.0414	1					
CV_t	0.3113	0.1738	− 0.0490	0.0682	1				
FV_t	0.3803	0.2938	− 0.0451	0.0287	0.3944	1			
dCV_t	0.1665	0.0313	− 0.0023	0.0042	0.4905	0.1236	1		
dFV_t	0.2379	0.2451	− 0.0112	0.0023	0.1084	0.3912	0.2685	1	
Panel D: ADF statistics									
ADF Test 2	$p_{CL,t}$	$p_{FL,t}$	$\sigma_{CL,t}^2$	$\sigma_{FL,t}^2$	$\sigma_{FO,t}^2$	$\sigma_{IO,t}^2$	CV_t	FV_t	
Test statistics	− 0.2464	− 1.9394	− 13.53 ^c	− 3.973 ^c	− 12.195 ^c	− 10.14 ^c	− 1.49	− 1.89	
Max lag	1	18	13	13	15	9	12	13	

^aIf the Autocorrelation coefficient (Partial Autocorrelation Function or PACF) lies within the 95% confidence interval range, it is not significant. The 95% confidence interval for the PACF is (− 0.0247, 0.0247)

^bQ(k) is the Ljung-Box test-statistics of joint significance for first to kth order Autocorrelation, and is χ^2 -distributed

^cIndicates significant at 5% level

in the residuals up to the second order, which generates inefficient coefficient estimates. Accordingly, the estimation of ESVL(CI) involves a two-step procedure. First, estimate ESVL(CI) according to Eq. 6. This generates a series of residuals $u_{i,t}$ for each of the four equations. Next, include $u_{i,t-1}$ and $u_{i,t-2}$ in their corresponding equations as additional variables to proxy for other variables that are not explicitly considered. Table 3 reports an improvement in model fitting from an increase in the log-likelihood function.

Second, we test where the order in which volatility variables enter the various equations of the system based on our assumed informational pecking-order is appropriate. We benchmark each of the two ESVL models against two corresponding ESVL systems whether the order of efficiency is intentionally specified to be in stark contrast with ESLV(CI) and ESVL(FI). Table 4 reveals a huge difference in the log-likelihood functions e.g. the log-likelihood function for ESVL(CI) is 2138.01, whereas for its mis-specified counterpart, the value dropped to 1209.78. This suggests the importance of specifying the ESVL according to an appropriate order of volatility transmission.

ESVL(CI) and ESVL(FI) estimation results are reported in Tables 5 and 6 respectively. For ESVL(CI), $\sigma_{FI,t}$ is significant in the $\sigma_{CI,t}$ equation. Dum^{Open} is significant in all except the $\sigma_{FO,t}$ equation. Lastly, all lag-1 volatility variables are significant in their corresponding equations.

Cross-market volatility transmission effects are more evident in ESVL(FI). Both $\sigma_{CI,t}$ and $\sigma_{FO,t}$ are significant in the $\sigma_{FI,t}$ equation. Both $\sigma_{FI,t}$ and $\sigma_{IO,t}$ are significant in the $\sigma_{CI,t}$ equation. Contemporaneous and lagged volatility from both CI and FI are

Table 3 Log-likelihood function of the ESVL models and their augmented counterparts

	ESVL(CI) with lagged residuals	ESVL (CI)	ESVL(FI) with lagged residuals	ESVL (FI)
Log-likelihood function	2,138.01	2,094.51	1,872.68	1,647.87

Note: Strictly speaking, evaluation on improved model fitting where competing models are of different dimension has to be made based on other information criteria like the Schwartz-Bayesian Information Criteria (SBIC), or the Akaike Information Criterion (AIC). However, the vast improvement in parameter significance and the removal of auto-correlation in the disturbance terms of the augmented ESVL models are adequate justification to favor applying the augmented versions of the ESVL models

Table 4 Log-likelihood of ESVL models and mis-specified-ESVL models

	ESVL (CI)	Mis-specified ESVL (CI)	ESVL (FI)	Mis-specified ESVL (FI)
Log-likelihood function	2,138.01	1,209.78	1,872.68	1,650.41

Note: Here, model evaluation can be made based solely on the log-likelihood function since the corresponding competing models are of identical dimensions

Table 5 Results from the FIML estimation on the ESVL(CI)

	$\sigma_{CI, t}$	$\sigma_{FI, t}$	$\sigma_{IO, t}$	$\sigma_{FO, t}$
Constant	− 0.0751 (0.150) ^a	0.0581 (0.003) ^b	0.0157 (0.340)	− 2.9534 (0.809)
Dum^{open}	− 2.3323 (0.007) ^b	1.3919 (0.000) ^b	− 0.0162 (0.008) ^b	− 102.173 (0.754)
$\sigma_{CI, t}$	~	0.3875 (0.159)	0.0418 (0.8580)	− 48.1824 (0.739)
$\sigma_{FI, t}$	1.6797 (0.007) ^b	~	~	73.5818 (0.754)
$\sigma_{IO, t}$	0.01124 (0.535)	~	~	~
$\sigma_{FO, t}$	~	− 0.0002 (0.966)	− 0.0013 (0.874)	~
$\sigma_{CI, t-1}$	− 0.0691 (0.378)	0.0582 (0.061)	− 0.0104 (0.780)	− 1.1531 (0.948)
$\sigma_{FI, t-1}$	− 0.0012 (0.909)	0.0034 (0.587)	~	− 0.0627 (0.969)
$\sigma_{IO, t-1}$	~	~	0.8584 (0.000) ^b	~
$\sigma_{FO, t-1}$	~	0.0001 (0.983)	0.0011 (0.886)	0.9183 (0.000) ^b
CV_{t-1}	− 0.0001 (0.310)	~	~	~
$u_{CI, t-1}$	− 0.0069 (0.880)	~	~	~
$u_{CI, t-2}$	0.0484 (0.228)	~	~	~
$u_{FI, t-1}$	~	0.0157 (0.707)	~	~
$u_{FI, t-2}$	~	0.0380 (0.247)	~	~
$u_{IO, t-1}$	~	~	− 0.0238 (0.054)	~
$u_{IO, t-2}$	~	~	0.0364 (0.043) ^c	~
$u_{FO, t-1}$	~	~	~	− 0.2283 (0.000) ^b
$u_{FO, t-2}$	~	~	~	− 0.0196 (0.702)

^ap-values in parentheses^bSignificant at 1% level^cSignificant at 5% level

significant in the $\sigma_{FO, t}$ equation. This may be indicative of volatility feedback from the underlying asset to the futures index, and then from there back to the futures option market. Lagged residuals are all significant in their corresponding equations.

Table 6 Results from the FIML estimation on the ESVL(FI)

	$\sigma_{CI, t}$	$\sigma_{FI, t}$	$\sigma_{IO, t}$	$\sigma_{FO, t}$
Constant	0.4352 (0.000) ^{a,b}	-162.222 (0.027) ^c	-0.0114 (0.995)	-133.240 (0.000) ^b
Dum^{close}	0.3031 (0.000) ^b	-50.0409 (0.287)	0.1591 (0.982)	-82.7965 (0.000) ^b
$\sigma_{CI, t}$	~	129.865 (0.023) ^c	-0.0610 (0.875)	4.1211 (0.038) ^c
$\sigma_{FI, t}$	-0.6375 (0.000) ^b	~	~	1740.69 (0.000) ^b
$\sigma_{IO, t}$	~	10646.3 (0.006) ^b	~	~
$\sigma_{FO, t}$	0.1646 (0.019) ^c	~	0.3556 (0.088)	~
$\sigma_{CI, t-1}$	0.1908 (0.001) ^b	156.525 (0.053)	-0.1647 (0.840)	-40.9111 (0.001) ^b
$\sigma_{FI, t-1}$	~	63.9502 (0.842)	~	-118.136 (0.000) ^b
$\sigma_{IO, t-1}$	~	-9762.23 (0.004) ^b	0.9682 (0.000) ^b	~
$\sigma_{FO, t-1}$	-0.1659 (0.127)	~	-0.2979 (0.102)	1.5648 (0.644)
FV_{t-1}	-4029 (0.567)	~	~	~
$u_{CI, t-1}$	378.8470 (0.000) ^b	~	~	~
$u_{CI, t-2}$	-203.605 (0.000) ^b	~	~	~
$u_{FI, t-1}$	~	1.2902 (0.003) ^b	~	~
$u_{FI, t-2}$	~	0.2994 (0.022) ^c	~	~
$u_{IO, t-1}$	~	~	-0.2888 (0.000) ^b	~
$u_{IO, t-2}$	~	~	-0.1597 (0.002) ^b	~
$u_{FO, t-1}$	~	~	~	-0.9268 (0.000) ^b
$u_{FO, t-2}$	~	~	~	0.3531 (0.000) ^b

^ap-values in parentheses^bSignificant at 1% level^cSignificant at 5% level

Lastly, we report SVL estimation results in Table 7. All own-market lag-1 volatilities are significant. All lagged residuals are significant as well. Lagged futures tick volume is also significant in the $\sigma_{FI, t}$ equation. To note, $\sigma_{FI, t}$ is significant and negative in the $\sigma_{CI, FI, t}$ equation. This result is consistent with our argument of an

Table 7 Results from the FIML estimation on SVL

	$\sigma_{FI, t}$	$\sigma_{CI, t}$	$\sigma_{CI, FI, t}$
Constant	-7.5256 (0.000) ^{a,b}	-1.0675 (0.573)	-46.3706 (0.000) ^b
Dum^{close}	0.3108 (0.093)	-1.1309 (0.001) ^b	-0.7649 (0.000) ^b
$\sigma_{FI, t}$	~	-0.0414 (0.669)	-0.2503 (0.049) ^c
$\sigma_{CI, t}$	~	~	~
$\sigma_{FI, t-1}$	-1.3060 (0.000) ^b	~	~
$\sigma_{CI, t-1}$	~	1.0742 (0.000) ^b	0.3628 (0.000) ^b
$\sigma_{t-1}^{CI, FI}$	0.0667 (0.197)	0.2191 (0.153)	5.6177 (0.000) ^b
FV_{t-1}	0.0117 (0.005) ^b	~	~
$u_{FI, t-1}$	-0.4869 (0.001) ^b	~	~
$u_{FI, t-2}$	-0.0794 (0.000) ^b	~	~
$u_{CI, t-1}$	~	-1.1272 (0.000) ^b	~
$u_{CI, t-2}$	~	-0.1377 (0.000) ^b	~
$u_{CI, FI, t-1}$	~	~	-5.6253 (0.000) ^b
$u_{CI, FI, t-2}$	~	~	-0.8211 (0.000) ^b

^ap-values in parentheses^bSignificant at 1% level^cSignificant at 5% level

inverse relation between $\sigma_{FI, t}$ and $\sigma_{CI, FI, t}$ since the futures market experiences more trading activity than the underlying asset market in response to (say) unexpected macroeconomics news.

61.5 Out-of-Sample Hedging Performances

In this section, we evaluate the incremental profits from rebalancing the futures position that competing OHRs generate over a static hedge. Indeed, given that true volatility is unobservable, an economic evaluation in the context of an out-of-sample dynamic hedging scheme is more meaningful than a statistical evaluation. To reiterate, our focus on incremental profits from rebalancing the futures position rather than the standard risk minimization as our hedging performance criterion stems from the focus of our paper. We are interested in whether incremental

information in implied volatility translates into incremental profits from rebalancing the futures position.

61.5.1 Details of the Hedging Scheme

We consider hedging against the exposure of a long position in a widely-held index-tracking portfolio with short positions in index futures contracts. We refer to the combined cash and futures positions as our net position. A static hedge involves no interim re-balancing during the hedging period. In contrast, a dynamic hedge involves interim re-balancing of the futures position during the hedging period, thereby accumulating interim profits or losses. In our dynamic hedging scheme, we consider the following scenario. An institutional trader owns a \$five million (m) widely-held equity portfolio that is very highly correlated with the S&P 500 index.¹⁹ The objective is to protect the portfolio against market downside risk, and S&P 500 futures contracts is the only hedging instrument that will be considered.

The size of the futures position is also affected by $V_{CI,t}$ and $V_{FI,t}$. The latter is calculated as $V_{FI,t} = p_{FI,t} \times 500$ i.e. the time t futures price times the contract multiplier of \$500 per index point. Note that $V_{CI,t=0} = \$5$ m. Let $V_{CI,t=1} = 5m \times (1 + r_{CI,t=1})$ and $V_{CI,t=2} = V_{CI,t=1} \times (1 + r_{CI,t=2})$ etc. Accordingly, the half-hourly profit/loss from the physical portfolio at time t is $V_{CI,t} - V_{CI,t-1}$. The change in value of a futures contract at time t is $(p_{FI,t} - p_{FI,t-1}) \times 500$. Since these are short positions, a profit arises when $(p_{FI,t} - p_{FI,t-1}) < 0$. The interim profit from re-balancing the futures position at time t is calculated as $(V_{FI,t-1} - V_{FI,t}) \times (N_{t-1} - N_{t-2})$, where $(N_{t-1} - N_{t-2})$ is the 'Number of contracts rebalanced at time $t - 1$ '.²⁰ To elaborate, if $(N_{t-1} - N_{t-2}) = +4$, this implies that an additional four contracts were shorted at time $t - 1$. Subsequently, if the value of the futures contract decreases between time $t - 1$ and time t e.g. $V_{FI,t-1} - V_{FI,t} = +1,000$, then the interim profit at time t generated from the rebalancing at time $t - 1$ is \$4000. If $N_{t-1} - N_{t-2} = 0$, this implies no rebalancing is required at time $t - 1$, such that the interim profit for time t is zero.

Time series plots reveal that ESVL-based OHRs exhibit more intraday variability relative to both GARCH-based OHR and IHR. Accordingly, in the hedging exercise, we allow interim rebalancing as frequently as every half-hour, which implies that rebalancing will occur more often for ESVL-based OHRs. However, this does not necessarily imply that that ESVL-based OHRs will be more profitable given the presence of transaction costs and the fact that some rebalancing transactions will generate interim losses. Intraday rebalancing does occur for certain OHRs on some days during the test period. However, consecutive half-hourly rebalancing seldom occurs. As our test sample covers the second half of 1990, we conduct our hedging exercise separately on the

¹⁹ As mentioned in the introduction, we assume the S&P 500 index as our physical portfolio.

²⁰ Note that 'N', which represents the number of futures contracts, is always positive. The 'No. of contracts rebalanced at time $t - 1$ ' is defined as $(N_{t-1} - N_{t-2})$. As such, if $N_{t-1} = 5$ and $N_{t-2} = 3$, then the 'No. of contracts rebalanced at time $t - 1$ ' is +2, which implies shorting an additional two futures contracts.

September and December contracts. We assume that the institutional investor rolled over from the September to December contract after the eighth trading day in September 1990, which is consistent with our sampling procedure.

61.5.2 Incremental Profit Results

The incremental profits generated by competing OHRs are reported in Table 8. In the September quarter of 1990, the US equity market experienced a sharp downturn. This is followed by a moderate recovery in the December quarter. The value of the physical portfolio decreased by \$771,716 or 15.43% in the September quarter. While it gained back \$150,458 by the end of the December quarter, what transpired is an overall half-year loss of 12.43% on the initial investment value of \$5 m. The static hedge in OHR(1), which is ranked third overall, performed surprisingly well. By construction, OHR(1) is the only hedge ratio that counter-balances any profit (loss) from the physical portfolio. However, OHR(1) is unable to completely preserve the value of the physical portfolio, and resulted in an overall loss of \$165,858. However, this is still better than having an unhedged position.

Next, the table shows that OHR(5) and OHR(8), which are generated from ESVL(CI) and ESVL(FI), clearly outperform all other OHRs considered in this chapter. In addition to being the only two hedge ratios that outperform the static hedge OHR(1), these ESVL-based OHRs are also the only OHRs that manage to generate sufficient profits from interim rebalancing to cover the overall loss of \$621,258 incurred by the physical portfolio during the test period. The enhanced GARCH-based OHR(3) is ranked fourth overall. It performs well in that it is able

Table 8 Out-of-sample incremental profits generated by competing hedge ratios

OHR ranking		Profit from Sep contract	Profit from Dec contract	Profit from both contracts	Combined cash and futures position
1st place OHR(5)	Constant covariance ESVL(CI) and ESVL(FI)	\$1,591,700	\$145,450	\$1,737,150	\$1,115,892
2nd place OHR(8)	Time varying covariance SVL, ESVL(CI) and ESVL(FI)	\$1,545,225	\$150,100	\$1,695,325	\$1,074,067
3rd place OHR(1)	Static hedge	\$478,800	\$(23,400)	\$455,400	\$(165,858)
4th place OHR(3)	Enhanced GARCH (1,1)	\$52,325	\$15,775	\$68,100	\$(553,158)
5th place OHR(7)	Time varying covariance IHR	\$13,025	\$19,225	\$32,250	\$(589,008)
6th place OHR(4)	Constant covariance IHR	\$12,400	\$18,300	\$30,700	\$(590,558)
7th place OHR(6)	SVL	\$(39,175)	\$59,300	\$20,125	\$(601,133)
8th place OHR(2)	GARCH (1,1)	\$(239,425)	\$(73,625)	\$(313,050)	\$(934,308)
Profit/loss from the physical portfolio		\$(771,716)	\$150,458	\$(621,258)	\$(621,258)

to generate interim profits of \$52,325 and \$15,775 from both the September and December contracts, although its overall profit of \$68,100 is lower than that of the static hedge. Despite incorporating additional volatility variables in the weighting series, OHR(3) is overshadowed by the ESVL-based OHRs. Similar comments apply to OHR(4) and OHR(7), the two OHRs based on implied volatilities from the IO and FO markets.

OHR(6), which focuses on modeling time-varying covariance between the spot and futures prices, performed poorly. It improves the unhedged position by a modest \$20,125. The latter consists of a loss of \$39,175 in the September contract and a profit of \$59,300 in the December contract. To note, the gain and loss in the futures position generated by OHR(6) are in the same direction as that of the underlying portfolio. This implies that OHR(6) exacerbates the variability of the overall position, rather than reducing it. OHR(2), which is based on the standard GARCH(1,1), is ranked last. It generates losses from both contracts, with an accumulated loss of \$313,050 in the futures position alone.

We draw two implications from the preceding results. First, the findings support our proposition that incorporating incremental information from implied volatilities to generate the OHR projections translates into incremental profits from interim balancing the futures position. OHRs based solely on cash-futures data or options data, have all under-perform OHRs derived from a combination of cash, futures and options data, with the exception of the static hedge. The enhanced GARCH-based OHR(3) is superior to OHR(4), which is based solely on implied volatilities. It also outperforms OHR(6) and OHR(2), although OHR(3) itself failed to outperform the static hedge. Second, the profit dominance of ESVL-based OHR(5) and OHR(8) over OHR(3) suggests that the ESVL framework is more suitable than GARCH (1,1) at modeling intraday volatility transmission across the four S&P markets for the purpose of making out-of-sample OHR projections.

In general, the consideration of transaction costs offers a realistic balance of the incremental costs and benefits from interim rebalancing when bench-marked against a static hedge. However, we argue that transaction cost is a moot consideration for our paper. This is because most of the OHRs we consider, including GARCH, SVL and implied volatility, are already ranked below the static hedge OHR(1) based on gross profit. On the other hand, the ESVL-based OHR(5) and OHR(8) do instigate very frequent re-balancing of large numbers of futures contracts throughout the test period. It is obvious that any interim rebalancing based on either OHR(5) or OHR(8) will accumulate the most transaction costs. However, note the extravagant incremental profits from OHR(5) and OHR(8) of \$1,591,700 and \$1,545,225 correspondingly from rebalancing the September contract, and \$145,450 and \$150,100 from rebalancing the December contract. For the combined September and December futures positions, each of OHR(5) and OHR(8) generated interim profits that is approximately 2.5 times the loss incurred by the physical portfolio. The presence of transaction cost is unlikely to alter the extreme dominance of the ESVL-based OHRs over the static OHR(1).

61.6 Conclusion

If option markets contribute incremental information to their underlying asset markets, then incorporating implied volatility into modeling and forecasting dynamic hedge ratios should translate into incremental profits from interim rebalancing. Our results support this proposition. In addition, we show that such cross-market volatility transmission effects are captured by a system of simultaneous volatility equations. Hedge ratios generated from extended volatility systems ESVL, which incorporate intraday volatility transmissions across all four S&P 500 cash, futures, index option and futures option markets, significantly outperform all other competing OHRs considered in this chapter.

Furthermore, the ESVL-based OHRs are the only hedge ratios that manage to generate profit from rebalancing the futures positions across both the September and December contracts in excess of the losses incurred by the physical portfolio. In fact, the extravagant profit results reported in Table 8 should attract considerable attention not only from hedgers. Since institutional speculators do not hold the physical portfolio, their focus is solely on the profit/loss generated from actively rebalancing the futures position based on a time-varying measure of the cash-futures sensitivity. The out-of-sample OHR projections can be easily applied in a trading strategy using index futures.

61.7 Deriving the Reduced-Form of ESVL(CI)

The appendix provides details on a time-series of $\sigma_{CI, t+1}$ is generated from Eq. 6 for constructing hedge ratios. To note, a similar methodology applies for generating a time-series of $\sigma_{FI, t+1}$. The key step involves deriving the reduced-form $\sigma_{CI, t}$ equation corresponding to Eq. 6. We show that the reduced-form $\sigma_{CI, t}$ equation contains only lagged volatility variables, which is required to be able to generate one-step ahead forecasts of $\sigma_{CI, t+1}$. The coefficients of the reduced-form $\sigma_{CI, t}$ are functions of coefficients from the structural system. These coefficients are derived by fitting the structural system with sample over the estimation period. These, together with time t volatility variables, allows us to obtain $\sigma_{CI, t+1}$. After each forecast, the estimation period is expanded by 1 observation; the coefficients are updated and $\sigma_{CI, t+1}$ is calculated as per normal. This recursive process generates a time-series of $\sigma_{CI, t+1}$. To note, while generating such hedge ratios may seem computationally tedious, but since the projections are made one-step at a time, the coefficients are stable even as we sequentially move across the test period. Thus updating the coefficient estimates is fairly easy.

The following describes the substitution process based on Eq. 6 to obtain the reduced-form $\sigma_{CI, t}$. The discussion emphasizes only on the contemporaneous

volatility variables. Denote the individual equations in the system below as A1, A2, A3 and A4. A1 to A7 demonstrates how we derive from the four-equation structural system to the reduced form system. The CI and FI reduced-form equations are then used to make out-of-sample volatility forecasts.

$$\begin{aligned}
 \sigma_{CI,t} &= \alpha_{10} + \alpha_{11}Dum^{Open} + \beta_{11}(\sigma_{FI,t}) + \beta_{12}(\sigma_{IO,t}) \\
 &\quad + \gamma_{11}(\sigma_{CI,t-1}) + \gamma_{12}(\sigma_{FI,t-1}) + \gamma_{13}(CV_{t-1}) + \varepsilon_{CI,t} \\
 \sigma_{FI,t} &= \alpha_{20} + \alpha_{21}Dum^{Open} + \beta_{21}(\sigma_{CI,t}) + \beta_{22}(\sigma_{FO,t}) \\
 &\quad + \gamma_{21}(\sigma_{CI,t-1}) + \gamma_{22}(\sigma_{FI,t-1}) + \gamma_{23}(\sigma_{FO,t-1}) + \varepsilon_{FI,t} \\
 \sigma_{IO,t} &= \alpha_{30} + \alpha_{31}Dum^{Open} + \beta_{31}(\sigma_{CI,t}) + \beta_{32}(\sigma_{FO,t}) \\
 &\quad + \gamma_{31}(\sigma_{CI,t-1}) + \gamma_{32}(\sigma_{IO,t-1}) + \gamma_{33}(\sigma_{FO,t-1}) + \varepsilon_{IO,t} \\
 \sigma_{FO,t} &= \alpha_{40} + \alpha_{41}Dum^{Open} + \beta_{41}(\sigma_{CI,t}) + \beta_{42}(\sigma_{FI,t}) \\
 &\quad + \gamma_{41}(\sigma_{CI,t-1}) + \gamma_{42}(\sigma_{FI,t-1}) + \gamma_{43}(\sigma_{FO,t-1}) + \varepsilon_{FO,t}
 \end{aligned}$$

First, note that the $\sigma_{FO,t}$ equation contains $(\sigma_{CI,t}, \sigma_{FI,t})$ and the $\sigma_{FI,t}$ equation contains $(\sigma_{CI,t}, \sigma_{FO,t})$. Accordingly, substituting $\sigma_{FO,t}$ into $\sigma_{FI,t}$ allows us to express $\sigma_{FI,t}$ in terms of $\sigma_{CI,t}$. Second, substitute $\sigma_{FI,t}$ in terms of $\sigma_{CI,t}$ into $\sigma_{FO,t}$, such that $\sigma_{FO,t}$ can also be expressed entirely in terms of $\sigma_{CI,t}$. Third, substitute $\sigma_{FO,t}$ in terms of $\sigma_{CI,t}$ into $\sigma_{IO,t}$ such that $\sigma_{IO,t}$ can also be expressed entirely in terms of $\sigma_{CI,t}$. Lastly, substitute $\sigma_{FI,t}$ and $\sigma_{IO,t}$ in terms of $\sigma_{CI,t}$ into $\sigma_{CI,t}$ thereby obtaining the reduced-form $\sigma_{CI,t}$ which is expressed entirely in terms of lagged volatility variables.

From equation A4:

$$\begin{aligned}
 \beta_{22}\sigma_{FO,t} &= \beta_{22}\beta_{41}\sigma_{CI,t} + \beta_{22}\beta_{42}\sigma_{FI,t} \\
 &\quad + \beta_{22}[\alpha_{40} + \alpha_{41}Dum^{Open} + \gamma_{41}(\sigma_{CI,t-1}) \\
 &\quad + \gamma_{42}(\sigma_{FI,t-1}) + \gamma_{43}(\sigma_{FO,t-1}) + \varepsilon_{FO,t}]
 \end{aligned}$$

Substitute preceding into equation A2 and label as equation A5:

$$\begin{aligned}
 \sigma_{FI,t} &= \alpha_{20} + \alpha_{21}Dum^{Open} + \beta_{21}(\sigma_{CI,t}) + \beta_{22}\beta_{41}\sigma_{CI,t} \\
 &\quad + \beta_{22}\beta_{42}\sigma_{FI,t} + \beta_{22}[\alpha_{40} + \alpha_{41}Dum^{Open} + \gamma_{41}(\sigma_{CI,t-1}) \\
 &\quad + \gamma_{42}(\sigma_{FI,t-1}) + \gamma_{43}(\sigma_{FO,t-1}) + \varepsilon_{FO,t}] \\
 &\quad + \gamma_{21}(\sigma_{CI,t-1}) + \gamma_{22}(\sigma_{FI,t-1}) + \gamma_{23}(\sigma_{FO,t-1}) + \varepsilon_{FI,t} \\
 \sigma_{FI,t} &= \frac{\beta_{21} + \beta_{22}\beta_{41}}{1 - \beta_{22}\beta_{42}}\sigma_{CI,t} + \frac{1}{1 - \beta_{22}\beta_{42}}[\alpha_{20} + \alpha_{21}Dum^{Open} \\
 &\quad + (\gamma_{21} + \beta_{22}\gamma_{41})\sigma_{CI,t-1} + (\gamma_{22} + \beta_{22}\gamma_{41})\sigma_{FI,t-1} \\
 &\quad + (\gamma_{23} + \beta_{22}\gamma_{43})\sigma_{FO,t-1} + \beta_{22}\varepsilon_{FO,t} + \varepsilon_{FI,t}]
 \end{aligned}$$

Substitute equation A5 into A4 and label as equation A6:

$$\begin{aligned}
 \sigma_{FO,t} &= \beta_{41}\sigma_{CI,t} + \frac{\beta_{42}}{1 - \beta_{22}\beta_{42}} [(\beta_{21} + \beta_{22}\beta_{42})\sigma_{CI,t} \\
 &\quad + (\alpha_{20} + \beta_{22}\alpha_{40}) + (\alpha_{21} + \beta_{22}\alpha_{41})Dum^{Open} \\
 &\quad + (\gamma_{21} + \beta_{22}\gamma_{41})\sigma_{CI,t-1} + (\gamma_{22} + \beta_{22}\gamma_{42})\sigma_{FI,t-1} \\
 &\quad + (\gamma_{23} + \beta_{22}\gamma_{43})\sigma_{FO,t-1} + \beta_{22}\varepsilon_{FO,t} + \varepsilon_{FI,t}] + \alpha_{40} \\
 &\quad + \alpha_{41}Dum^{Open} + \gamma_{41}\sigma_{CI,t-1} + \gamma_{42}\sigma_{FI,t-1} \\
 &\quad + \gamma_{43}\sigma_{FO,t-1} + \varepsilon_{FO,t} \\
 \sigma_{FO,t} &= \frac{\beta_{41} + \beta_{21}\beta_{22}}{1 - \beta_{22}\beta_{42}}\sigma_{CI,t} + \frac{1}{1 - \beta_{22}\beta_{42}} \\
 &\quad \times [(\alpha_{40} + \alpha_{20}\beta_{42})Dum^{Open} + (\gamma_{41} + \gamma_{21}\beta_{42})\sigma_{CI,t-1} \\
 &\quad + (\gamma_{42} + \gamma_{22}\beta_{42})\sigma_{FI,t-1} + (\gamma_{43} + \gamma_{23}\beta_{42})\sigma_{FO,t-1} \\
 &\quad + \beta_{42}\varepsilon_{FI,t} + \varepsilon_{FO,t}]
 \end{aligned}$$

Substitute equation A6 into A3 and label as equation A7:

$$\begin{aligned}
 \sigma_{IO,t} &= \beta_{31}\sigma_{CI,t} + \beta_{32}\frac{\beta_{41} + \beta_{21}\beta_{22}}{1 - \beta_{22}\beta_{42}}\sigma_{CI,t} \\
 &\quad + \frac{\beta_{32}}{1 - \beta_{22}\beta_{42}} [(\alpha_{40} + \alpha_{20}\beta_{42})Dum^{Open} \\
 &\quad + (\gamma_{41} + \gamma_{21}\beta_{42})\sigma_{CI,t-1} + (\gamma_{42} + \gamma_{22}\beta_{42})\sigma_{FI,t-1} \\
 &\quad + (\gamma_{43} + \gamma_{23}\beta_{42})\sigma_{FO,t-1} + \beta_{42}\varepsilon_{FI,t} + \varepsilon_{FO,t}] + \alpha_{30} \\
 &\quad + \alpha_{31}Dum^{Open} + \gamma_{31}(\sigma_{CI,t-1}) + \gamma_{32}(\sigma_{FI,t-1}) \\
 &\quad + \gamma_{33}(\sigma_{FO,t-1}) + \varepsilon_{IO,t} \\
 \sigma_{IO,t} &= \frac{\beta_{31} + \beta_{32}\beta_{41} + \beta_{21}\beta_{22}\beta_{32} - \beta_{22}\beta_{31}\beta_{42}}{1 - \beta_{22}}\sigma_{CI,t} \\
 &\quad + \frac{1}{1 - \beta_{22}\beta_{42}} [(\beta_{30} + \beta_{40}\beta_{32} + \alpha_{20}\beta_{32}\beta_{42} - \alpha_{30}\beta_{22}\beta_{42}) \\
 &\quad + (\alpha_{31} + \alpha_{41}\beta_{32} + \alpha_{21}\beta_{32}\beta_{42} - \gamma_{31}\beta_{22}\beta_{42}) \\
 &\quad (Dum^{Open} + \sigma_{CI,t-1}) \\
 &\quad + (\alpha_{32} + \gamma_{42}\beta_{32} + \gamma_{22}\beta_{32}\beta_{42} - \gamma_{32}\beta_{22}\beta_{42})\sigma_{FI,t-1} \\
 &\quad + (\gamma_{33} + \gamma_{43}\beta_{32} + \gamma_{23}\beta_{32}\beta_{42} - \gamma_{33}\beta_{22}\beta_{42})(\sigma_{FO,t-1}) \\
 &\quad + \beta_{32}(\varepsilon_{FO,t-1} + \beta_{42}\varepsilon_{FI,t-1}) + \varepsilon_{IO,t}]
 \end{aligned}$$

Substitute equations A5 and A7 into A1 to obtain the reduced-form.

$$\begin{aligned}
 \sigma_{CI,t} = & \beta_{11} \frac{\beta_{21} + \beta_{22}\beta_{41}}{1 - \beta_{22}\beta_{42}} \sigma_{CI,t} + \frac{\beta_{11}}{1 - \beta_{22}\beta_{42}} [(\alpha_{20} + \alpha_{40}\beta_{22}) \\
 & + (\alpha_{21} + \alpha_{41}\beta_{22})Dum^{Open} \\
 & + (\gamma_{21} + \beta_{22}\gamma_{41})\sigma_{CI,t-1} + (\gamma_{22} + \beta_{22}\gamma_{42})\sigma_{FI,t-1} \\
 & + (\gamma_{23} + \beta_{22}\gamma_{43})\sigma_{FO,t-1} + \beta_{22}\varepsilon_{FO,t} + \varepsilon_{FI,t}] \\
 & + \beta_{12} \frac{\beta_{31} + \beta_{32}\beta_{41} + \beta_{21}\beta_{22}\beta_{32} - \beta_{22}\beta_{31}\beta_{42}}{1 - \beta_{22}\beta_{42}} \sigma_{CI,t} \\
 & + \frac{\beta_{12}}{1 - \beta_{22}\beta_{42}} [(\alpha_{30} + \alpha_{40}\beta_{32} + \alpha_{20}\beta_{32}\beta_{42} - \alpha_{30}\beta_{22}\beta_{42}) \\
 & + (\alpha_{31} + \alpha_{41}\beta_{32} \\
 & + \alpha_{21}\beta_{32}\beta_{42} - \gamma_{31}\beta_{22}\beta_{42})Dum^{Open} \\
 & + (\alpha_{31} + \alpha_{40}\beta_{32} + \alpha_{21}\beta_{32}\beta_{42} - \gamma_{31}\beta_{22}\beta_{42})\sigma_{CI,t-1} \\
 & + (\alpha_{32} + \gamma_{42}\beta_{32} + \gamma_{22}\beta_{32}\beta_{42} - \gamma_{32}\beta_{22}\beta_{42})\sigma_{FI,t-1} \\
 & + (\gamma_{33} + \gamma_{43}\beta_{32} + \gamma_{23}\beta_{32}\beta_{42} - \gamma_{33}\beta_{22}\beta_{42})\sigma_{FO,t-1} \\
 & + \beta_{32}(\varepsilon_{FO,t} + \beta_{42}\varepsilon_{FI,t}) + \varepsilon_{IO,t}] \\
 \sigma_{CI,t} = & \frac{\beta_{11}}{\mathbf{A}} [(\alpha_{20} + \alpha_{40}\beta_{21}) + (\alpha_{21} + \alpha_{41}\beta_{22})Dum^{Open} \\
 & + (\gamma_{21} + \gamma_{41}\beta_{22})\sigma_{CI,t-1} + (\gamma_{22} + \gamma_{42}\beta_{22})\sigma_{FI,t-1} \\
 & + (\gamma_{23} + \gamma_{43}\beta_{22})\sigma_{FO,t-1} + \varepsilon_{FI,t}] \\
 & + \frac{\beta_{12}}{\mathbf{A}} [(\alpha_{30} + \alpha_{40}\beta_{32} + \alpha_{20}\beta_{32}\beta_{42} - \alpha_{30}\beta_{22}\beta_{42}) \\
 & + (\alpha_{31} + \alpha_{41}\beta_{32} + \alpha_{21}\beta_{32}\beta_{42} - \alpha_{31}\beta_{22}\beta_{42})Dum^{Open} \\
 & + (\alpha_{31} + \alpha_{41}\beta_{32} + \alpha_{21}\beta_{32}\beta_{42} - \alpha_{31}\beta_{22}\beta_{42})\sigma_{CI,t-1} \\
 & + (\alpha_{32} + \gamma_{42}\beta_{32} + \gamma_{22}\beta_{32}\beta_{42} - \gamma_{32}\beta_{22}\beta_{42})\sigma_{FI,t-1} \\
 & + (\gamma_{33} + \gamma_{43}\beta_{32} + \gamma_{23}\beta_{32}\beta_{42} - \gamma_{33}\beta_{22}\beta_{42})\sigma_{FO,t-1} \\
 & + \beta_{32}(\varepsilon_{FO,t} + \beta_{42}\varepsilon_{FI,t}) + \varepsilon_{IO,t}]
 \end{aligned}$$

where

$$\mathbf{A} = (1 - \beta_{22}\beta_{42} + \beta_{11}\beta_{21} + \beta_{11}\beta_{22}\beta_{41} + \beta_{12}\beta_{31} + \beta_{12}\beta_{32}\beta_{41} + \beta_{12}\beta_{21}\beta_{22}\beta_{32} - \beta_{12}\beta_{22}\beta_{31}\beta_{42})$$

The preceding demonstrates that Eq. 6 can be ‘collapsed’ to express $\sigma_{CI,t}$ in terms of lag-1 volatility variables.

References

- Au-Yeung, S., and G. Gannon. 2005. Regulatory change and structural effects in HSIF and HSI volatility. *Review of Futures Markets* 14: 283–308.

- . 2008. Modelling regulatory change v.s. volume of trading effects in HSIF and HSI volatility. *Review of Pacific Basin Financial Markets and Policies* 11: 47–59.
- Baillie, R., and R.J. Myers. 1991. Bivariate GARCH estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics* 6: 109–124.
- Bhattacharya, P., H. Singh, and G. Gannon. 2007. Time varying hedge ratios: An application to the Indian stock futures market. *Review of Futures Markets* 16: 75–104.
- Black, F. 1976. The pricing of commodity contracts. *Journal of Financial Economics* 3: 167–179.
- Chng, M., and G. Gannon. 2003. Contemporaneous intraday volume, options and futures volatility transmission across parallel markets. *International Review of Financial Analyst* 12: 49–68.
- Ederington, L. 1979. The hedging performance of the new futures markets. *Journal of Finance* 34: 157170.
- Engle, R. 2002. Dynamic conditional correlation: A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20: 339–350.
- Engle, R., and K. Kroner. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* 11: 122–150.
- Gannon, G. 1994. Simultaneous volatility effects in index futures. *Review of Futures Markets* 13: 1027–1066.
- . 2005. Simultaneous volatility transmissions and spillover effects: U.S. and Hong Kong stock and futures markets. *International Review of Financial Analysis* 14: 326–336.
- . 2010. Simultaneous volatility transmissions and spillover effects. *Review of Pacific Basin Financial Markets and Policies* 13: 127–156.
- Howard, C., and L. D'Antonio. 1984. A risk-return measure of hedging effectiveness. *Journal of Financial and Quantitative Analysis* 19: 101–112.
- Hsin, C., J. Kuo, and C. Lee. 1994. A new measure to compare the hedging effectiveness of foreign currency futures versus options. *Journal of Futures Markets* 14: 685707.
- Kroner, K., and J. Sultan. 1993. Time varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis* 28: 535–551.
- Lee, C., F. Lin, and M. Chen. 2010. International hedge ratios for index futures market: A simultaneous equations approach. *Review of Pacific Basin Financial Markets and Policies* 13: 203–213.
- Lee, C., K. Wang, and Y. Chen. 2009. Hedging and optimal hedge ratios for international index futures markets. *Review of Pacific Basin Financial Markets and Policies* 12: 593–610.
- Lee, W., G. Gannon, and C. Yeh. 2000. Risk and return performance of dynamic hedging models in international markets. *Asia Pacific Journal of Finance* 3: 125–148.
- Miller, M., J. Muthuswamy, and R. Whaley. 1994. Mean reversion of Standard & Poor's 500 Index basis changes: Arbitrage-induced or statistical illusion? *Journal of Finance*: 479–513.
- Park, T.H., and L. Switzer. 1995. Time varying distributions and the optimal hedge ratios for stock index futures. *Applied Financial Economics* 5: 131–137.
- Pennings, J., and M. Meulenberg. 1997. Hedging efficiency: A futures exchange management approach. *Journal of Futures Markets* 17: 599–615.
- Tse, Y., and A. Tsui. 2002. A multivariate generalized autoregressive conditional heteroscedasticity model with time varying correlations. *Journal of Business and Economic Statistics* 20: 351–362.
- Whaley, R. 1986. Valuation of American futures options: Theory and empirical tests. *Journal of Finance* 41: 127–150.
- Yeh, S., and G. Gannon. 2000. Comparing trading performance of the constant and dynamic hedge models: A note. *Review of Quantitative Finance and Accounting* 14: 155–160.



Lan-chih Ho, John Cadle, and Michael Theobald

Abstract

A portfolio insurance strategy is a dynamic hedging process that provides the investor with the potential to limit downside risk while allowing participation on the upside so as to maximize the terminal value of a portfolio over a given investment horizon. This chapter firstly introduces the basic concepts and payoffs of a portfolio insurance strategy. Secondly, it describes the theory of alternative portfolio insurance strategies. Thirdly, it provides the market developments of portfolio insurance strategies and real examples of structured products. Fourthly, it addresses the implications of these strategies on the financial market stability. Finally, it empirically compares the performances of various portfolio insurance strategies during different markets and time periods.

Keywords

CPPI · Expected shortfall · OBPI · Value-at-risk

L.-c. Ho (✉)

Central Bank of the Republic of China (Taiwan), Taipei, Taiwan, People's Republic of China

e-mail: LCHO@mail.cbc.gov.tw

J. Cadle · M. Theobald

University of Birmingham, Birmingham, UK

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,

https://doi.org/10.1007/978-3-030-91231-4_62

1437

62.1 Basic Concepts of Portfolio Insurance

Portfolio insurance refers to any strategy that protects the value of a portfolio of risky assets. The risky assets can be stocks, bonds, currencies, or even alternative assets, such as commodities, real assets, hedge funds, credits, etc. If the value of the risky asset declines, the insurance or hedge will increase in value to help offset the decline in price of the hedged risky assets. If the price of the risky asset increases, the increase of the insured portfolio will be less than the increase in the risky asset itself but will nevertheless still increase.

Table 1 illustrates how portfolio insurance works. In this example, the underlying risky asset is purchased for \$95 and \$5 is spent on portfolio insurance. The minimum amount that the insured investor can realize is \$95, but the uninsured portfolio can fall in value to a low of \$75 if the market falls. If the value of the risky asset increases, the value of the insured portfolio will increase, but at a smaller rate. Figure 1 illustrates the profit and loss of the insured and uninsured portfolio.

Portfolio insurance allows market participants to alter the return distribution to fit investors' needs and preferences for risk. Figure 2 shows the effect of insurance on the expected returns of a portfolio. Notice that the uninsured portfolio has greater upside potential as well as greater downside risk, whereas the insured portfolio limits the downside loss to the cost of the hedge. The upside potential of the insured portfolio is always below that of the uninsured portfolio. The cost of the insurance is the lower return for the insured portfolio should prices increase. While some investors would prefer the greater upside potential that the uninsured portfolio offers,

Table 1 Mechanics of portfolio insurance

Initial investment	\$100		
Cost of portfolio insurance	−\$ 5		
Amount of investment going toward securities	\$95		
Amount invested = \$100			
<i>Value of portfolio at year (dollars)</i>	<i>Return on uninsured portfolio (percent)</i>	<i>Value of insured portfolio (dollars)</i>	<i>Net return on insured portfolio (percent)</i>
75	−25	95	−5
80	−20	95	−5
85	−15	95	−5
90	−10	95	−5
95	−5	95	−5
100	0	95	−5
105	5	100	0
110	10	105	5
115	15	110	11
120	20	115	15
125	25	120	20
130	30	125	25

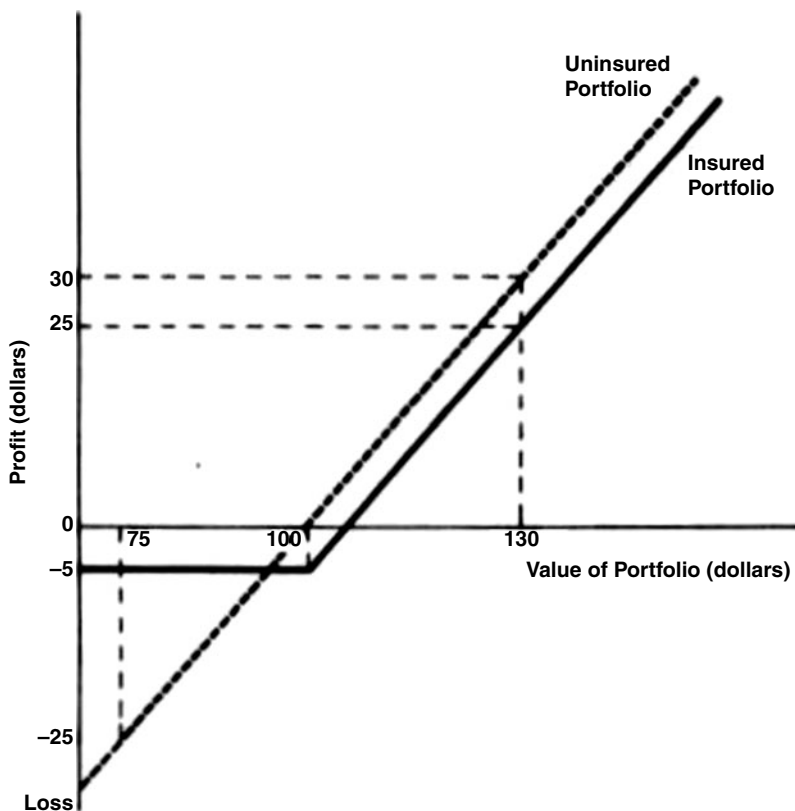


Fig. 1 Gains and losses of insured and uninsured portfolios

risk-averse investors would prefer the limited-risk characteristics that the hedged portfolio offers.

62.2 Theory of Alternative Portfolio Insurance Strategies

This section introduces the mechanisms behind various portfolio insurance strategies. These strategies share a common characteristic, a convex payoff function, which implies a “buy high and sell low” rule for the risky asset. If the market level declines, an investor sells a fraction of the risky asset and buys the riskless asset; if the market level rises, an investor switches from the riskless to the risky asset and rallies with the market. For illustrative purposes, we construct a portfolio that comprises a risky currency, the Euro (EUR), and a risk-free currency, the US dollar (USD) overnight deposit account. For comparison purposes, we assume that the initial value of EUR in the portfolio is 25%, short positions in EUR are not allowed, but that an investor can borrow via USD to buy EUR.

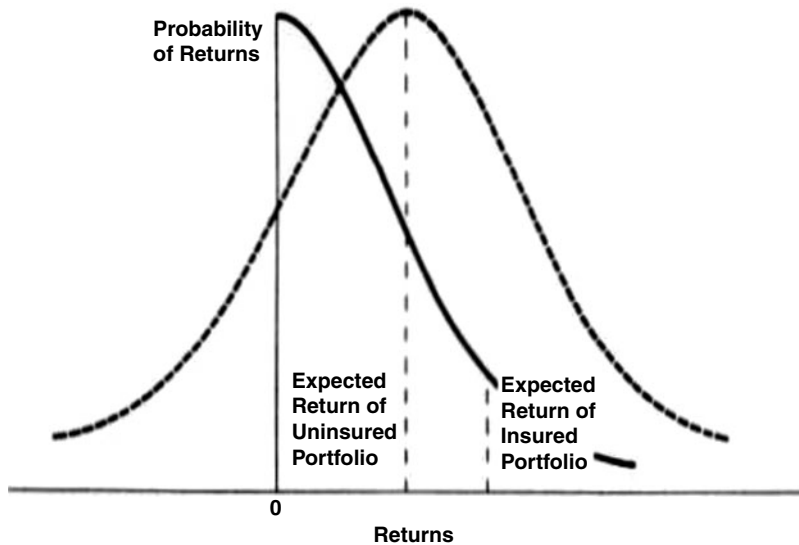


Fig. 2 Expected returns on insured and uninsured portfolios

62.2.1 Option-Based Portfolio Insurance (OBPI)

The most popular portfolio insurance strategy is the synthetic put approach of Rubinstein and Leland (1981), also referred to as an option-based portfolio insurance (OBPI) strategy. If an investor holds a risky asset and buys one at-the-money put option on that asset, the value of the resultant portfolio will not be less than the exercise price net of the premium at expiration. Thus, the investor effectively hedges the portfolio against downside risk. However, the implied volatility is usually higher than the historical volatility in option markets, which indicates that an actual option is more expensive than a synthetic one. Therefore, a synthetic option is often used as an alternative hedging vehicle in practice.

In a currency portfolio, a synthetic put can be created by dynamically rebalancing the portfolio between the risky (foreign) and riskless (domestic) currencies according to the delta of a put option,

$$D = e^{-rfT}[N(d_1) - 1] \quad (1)$$

Specifically, the investment proportion of the risky currency in the portfolio value can be expressed as:

$$W_{SynPut} = \frac{S + D * S}{V} = \frac{Se^{-rfT}N(d_1)}{S + P} \quad (2)$$

where P is the premium of a European put option on foreign currency from Garman-Kohlhagen (1983),

$$P = Xe^{-rT}N(-d_2) - Se^{-r_fT}N(-d_1) \quad (3)$$

and $d_1 = \frac{\ln(S/X) + (r - r_f + \sigma^2/2)T}{\sigma\sqrt{T}}$ and $d_2 = d_1 - \sigma\sqrt{T}$, S is the spot exchange rate, r is the domestic risk-free interest rate, r_f is the foreign risk-free interest rate, σ is the volatility of spot exchange rate return, X is the exercise price, T is the time to expiration, and $N(*)$ is the cumulative standard normal distribution function.

Assume that the initial portfolio contains 1,000,000 units of EUR. In order to make the initial value of the EUR position in the portfolio equal to 25%, an investor needs to create a synthetic put with a delta of -0.75 and sell 750,000 units of the EUR. The dollar income (\$990,075) from the EUR sold is kept in an USD cash account and earns the overnight deposit rate. Table 2 illustrates the adjustment process. As the EUR appreciates, the synthetic put moves less in-the-money with a delta of -0.7296 and an investor withdraws money from the USD cash account to buy more (20,407 units) EUR. On the contrary; as the EUR depreciates, the synthetic put moves further in-the-money with a delta of -0.7649 and an investor sells more (35,288 units) EUR. The daily adjustment is determined by the changes of delta.

One important parameter in the synthetic put approach is the volatility. Rendleman and O'Brien (1990), Do (2002), Do and Faff (2004), and Bertrand and Prigent (2005) use different measures of volatility, such as implied, historical and Leland's (1985) adjusted historical volatility, and discuss the effects of volatility estimation on the synthetic put strategy.

Additionally, in order to reduce transaction costs, some revisions on the rebalancing discipline are adopted, such as the time discipline with weekly/monthly rebalancing, the price discipline with adjustments according to changes in the value of the risky asset itself, or the volume discipline corresponding to changes in the number of shares that should be held. These revisions will be described in later section.

62.2.2 Constant Proportion Portfolio Insurance (CPPI)

A simplified approach, not involving complex "Greeks," is the constant proportion portfolio insurance (CPPI) strategy developed by Black and Jones (1987) and Perold (1986), and extended to fixed-income instruments by Hakanoglu et al. (1989). In CPPI, the exposure to the risky asset, E , is always kept at the cushion times a multiplier, m . The cushion, C , is the difference between the portfolio value, V , and a protected floor value, FL , and the multiplier is constant throughout the investment period:

$$E = m * C \quad (4)$$

$$C = V - FL \quad (5)$$

The investment proportion of the risky currency in the portfolio value can be expressed as:

Table 2 Daily adjustment process of portfolio insurance with synthetic put

Synthetic put	EURUSD S	Delta	EUR Amt sold (–)	EUR asset USD value	EUR asset %	\$ Income of EUR sold	USD Cum. income	Interest income	USD asset %	Portfolio USD value	Portfolio return%
		(1)	(2)	(3)	(4) = (3)/(9)	(5)	(6) = (5) + (6) + (7)	(7)	(8) = (7)/(9)	(9) = (3) + (6)	(10) = d (9)/(9)
2007/1/1	1.3201	–0.750000	–750,000	\$330,025	25.0%	\$990,075	\$990,075	\$145	75.0%	\$1,320,100	
2007/1/2	1.3272	–0.729593	20,407	\$358,884	27.1%	–\$27,084	\$963,137	\$142	72.9%	\$1,322,020	0.15%
2007/1/3	1.3170	–0.764881	–35,288	\$309,651	23.5%	\$46,474	\$1,009,753	\$148	76.5%	\$1,319,404	–0.20%

$$W_{CPPI} = \frac{E}{V} \quad (6)$$

In a portfolio of securities, the floor is usually set as the discount price of a zero-coupon bond, which approaches the par value at the end of the investment horizon. In a currency portfolio, the floor is set as the forward rate, with a maturity equal to the investment horizon, that appreciates with time, that is, $F_t = F_0 * \exp(r_0 * t)$, where $F_0 = S_0 * \exp((r_0 - r_{f,0}) * T)$. Note that when the domestic risk-free rate is higher than the foreign risk-free rate, the forward rate is at a premium, which would result in short-selling of the risky asset at the beginning of the investment horizon, that is, $E = m * C = m * (V - FL) = m * Units * (S - F) < 0$. In order to make the initial value of the EUR position in the portfolio equal to 25%, a discount factor (DF) is needed, and the multiplier¹ also needs to be trimmed at the beginning of investment horizon. Thus, the synthetic floor is set as:

$$F_t = DF * F_0 * \exp(r_0 * t) \quad (7)$$

Table 3 illustrates the adjustment process. As EUR appreciates, the portfolio value increases more than the floor level increases, the cushion increases, and investors buy more (5,238 units) EUR, vice versa. There are various market developments in CPPI mechanism with additional features, such as constraints on the investment level, constraints on leverage, variable floors, and variable multipliers, etc., which will be introduced in later section.

62.2.3 Risk-Based Portfolio Insurance (RBPI)

The idea of risk management has become widespread and has been applied to the asset allocation issue. By dynamically controlling for downside risk while maximizing the expected return of a portfolio, the concept of the modern risk-based asset allocation process is analogous to that of the classic portfolio insurance strategy. Zhao and Ziemba (2000) are the first to explore the portfolio payoffs between the two approaches. There are basically two popular measures of downside risk, one is the value-at-risk (VaR) and the other more coherent one is the expected shortfall (ES). Accordingly, there are two risk-based portfolio insurance (RBPI) strategies.

62.2.3.1 VaR-Based Portfolio Insurance

VaR-based portfolio insurance is a strategy that permanently controls the shortfall risk of the portfolio. The allocation to the risky asset is adjusted each day so that the shortfall probability – the probability of realizing a portfolio return (R_P) that falls

¹The magnitude of the multiplier depends on the discount factor. The lower the discount factor, the lower the multiplier in order to keep the initial exposure fixed at 25%, i.e., $E = m * (V - FL)$.

Table 3 Daily adjustment process of constant proportion portfolio insurance

CPPI	EURUSD S	Floor discount factor (DF) = 0.94; multiplier (m) = 5.24592					USD asset USD value	USD asset %	Portfolio return%
		Portfolio USD value	Floor level	EUR asset USD value	EUR amount	EUR asset %			
		(1)	(2)	(3) = (1)–(2) * m	(4)	(5) = (3)/(1)	(6) = (1)–(3)	(7) = (6)/ (1)	(8) = d(1)/ (1)
2007/1/1	1.3201	\$1,320,100	1.2572	\$330,025	250,000	25.0%	\$990,075	75.0%	
2007/1/2	1.3272	\$1,322,020	1.2574	\$338,752	255,238	25.6%	\$963,137	74.4%	0.15%
2007/1/3	1.3170	\$1,319,562	1.2577	\$324,510	246,401	24.6%	\$1,009,753	75.4%	–0.19%

below a pre-specified threshold (R) at the end of the investment period – does not exceed a target value, α (say 5%):

$$\text{Prob}(R_P < -R) \leq \alpha \quad (8)$$

To operationalize the shortfall risk, Herold et al. (2005) use the lower partial moment of order zero, which is equivalent to VaR, and assume a normally distributed portfolio returns:

$$LPM_0(R_P) = N\left(\frac{\tau - u}{\sigma}\right) \quad (9)$$

$$R_P = W * R_A + (1 - W) * R_F \quad (10)$$

where u and σ are the mean and the volatility of portfolio returns, respectively, and τ is the required portfolio return. R_A and R_F are the return of the risky and the riskless asset, respectively. Controlling for a fixed shortfall probability, the dynamic allocation process is solved for the weight of the risky asset so that the portfolio return is always higher than the required return (i.e., VaR):

$$NVaR_{\alpha,t}(R_P) = u_t(R_P) - N^{-1}(\alpha)\sigma_t(R_P) \quad (11)$$

Table 4 illustrates the daily adjustment process. Since the VaR of a portfolio return does not change dramatically at each day during a certain period of time, the frequency of adjustment is much less than that in traditional portfolio insurance strategies. However, as EUR depreciates further, the VaR of a portfolio return increases, and the allocation to the risky asset is suggested to decrease from 25% to 24.9%, an investor sells EUR.

In comparison with the traditional portfolio insurance strategy, Herold et al. (2005) show that the VaR-based model can be regarded as a generalized version of CPPI with a dynamic and time-varying implied multiplier. The risky currency investment proportion is then given by:

$$W_{VaR} = \frac{\tau}{-Z_\alpha * \sigma} \quad (12)$$

where Z_α is the α -quantile of the standard normal distribution and the process drift is assumed to be equal to zero at the daily differencing interval. The implied multiplier of the VaR model is obtained by:

$$m^{impl} = \frac{E}{C} = \frac{W * V}{C} = \frac{C/V}{Z_\alpha * \sigma} * \frac{V}{C} = \frac{1}{Z_\alpha * \sigma} = \frac{1}{VaR(R_P)} \quad (13)$$

where $\tau = -C/V$ is expressed in percentage terms. Thus, the inverse of the CPPI multiplier can be interpreted as the maximum loss or worst case return that is allowed to occur over the next period. Bertrand and Prigent (2001) propose that the upper bound on the multiplier can be determined by the extreme value theory.

Table 4 Daily adjustment process of VaR-based portfolio insurance

			$\alpha = 5\%$									
				</								

62.2.3.2 ES-Based Portfolio Insurance

ES-based portfolio insurance is a strategy that permanently controls the expected shortfall of the portfolio. The expected shortfall is defined as the expected value of the loss of a portfolio in a certain percentage of worst cases within a given holding period. The allocation to the risky asset is adjusted each day so that the expected shortfall does not exceed a target value.

Following the idea of Herold et al. (2005), Hamidi et al. (2009) propose a conditional CPPI multiplier that links the traditional CPPI strategy with the RBPI strategy. The conditional CPPI multiplier is determined by keeping the risk exposure of a portfolio constant, and the risk exposure can be defined either by the VaR or by the ES:

$$m^{cond} = \frac{1}{VaR_{\alpha}(R_P)} \quad \text{or} = \frac{1}{ES_{\alpha}(R_P)} \quad (14)$$

By assuming a normally distributed portfolio return, the expected shortfall and the VaR are scalar multiples of the standard deviation. There will be no differences across portfolio strategies employing variance, VaR, and ES risk measures. Therefore, Ho et al. (2011) use the historical distribution of portfolio returns to operate the ES-based portfolio insurance strategy:

$$HES_{\alpha,t}(R_P) = -E[R_P | R_P < -HV_{\alpha,t}(R_P)] \leq R \quad (15)$$

where

$HV_{\alpha,t}(R_P) = -\text{Percentile}(R_{P_{t-260}}, \dots, R_{P_t}; 5\%) \leq R$ among which $HV_{\alpha,t}(R_P)$ is the historical VaR of portfolio returns at date t at the $1-\alpha\%$ confidence level, $R_{P_{t-260}}, \dots, R_{P_t}$ indicates the daily portfolio return series given a certain weight of the risky asset during the past year, and $E[*]$ is the expectation operator.

Table 5 illustrates the daily adjustment process. Since the ES of a portfolio return does not change dramatically at each day during a certain period of time, the frequency of adjustment is also much less than that in the traditional portfolio insurance strategies.

As mentioned at the beginning of this section, all the portfolio insurance strategies have a “buy high and sell low” rule for the risky asset. It encounters a potential limit called “cash-lock” or “cash-out.” That is, if the market drops substantially an investor sells all fraction of the risky asset and locks in cash, the portfolio will never have a chance to rebound when market recovers. There are market variations to avoid this limit, which will be provided in later section.

62.3 Market Developments

In this section, the history of portfolio insurance strategy will be described, the recent market modifications in CPPI mechanisms and real examples of CPPI structured products will also be provided.

Table 5 Daily adjustment process of ES-based portfolio insurance

			$\alpha = 5\%$								
	EURUSD S	Portfolio USD value	EUR Adj. Rule	EUR amount	EUR asset USD value	EUR asset % %	USD deposit adjust (+)	Cumulated USD amount	Interest income	USD asset % %	Portfolio return%
		(1)	(2)	(3) = (2) * IniAmt	(4) = (3) * S	(5) = (4)/(1)	(6) = d (3) * S	(7) = (6) + (7) + (8)	(8)	(9) = (7)/(1)	(10) = d (1)/(1)
2007/1/1	1.3201	\$1,320,100	25.0%	250,000	\$330,025	25.0%		\$990,075	\$145.2	75.0%	
2007/1/2	1.3272	\$1,322,020	25.0%	250,000	\$331,800	25.1%	\$0	\$990,220	\$146.3	74.9%	0.15%
2007/1/3	1.3170	\$1,319,617	25.2%	251,613	\$331,374	25.1%	-\$2124	\$988,242	\$145.0	74.9%	-0.18%

62.3.1 History of Portfolio Insurance

Derivatives markets did not exist in 1970s, asset managers were required to use dynamic asset allocation techniques such as CPPI to preserve capital. Since the beginning of 1980, significant amount of the world equity portfolio were managed in accordance to CPPI (or OBPI) until the 1987 stock market crash, where the portfolio insurance technique was blamed to be a pro-cyclical mechanism.

In 1990s, the fast expanding derivatives markets made CPPI less interesting for asset managers, and CPPI was merely used in the structured products world as a capital protection methodology for hedge funds. At the end of 1990s, there was an increased usage of CPPI for equity and fixed income portfolios. However, the burst of Internet bubble in year 2000 resulted in many CPPIs going into cash-locked.

Since 2001, low interest rate environment and high implied volatility made portfolio insurance as a methodology for providing capital protection more interesting again. CPPI re-emerged due to low structuring and trading costs. Besides, nonconventional CPPI with advantageous modifications for investors expanded.

According to Pain and Rand (2008), OBPI investments have not been popular, which reflects the difficulty in explaining options to investors. The CPPI strategy, on the other hand, is much more prevalent and stemming from a broadening in the asset classes for which investors need principal protection. Many of principal-protected structured notes are not only designed to link the performance of traditional assets, such as stocks, currencies, interest rates, and bonds, but also invest in alternative risky assets, such as hedge funds, fund of hedge funds, credit derivatives, commodities, real estates, private equities, and any combination of the above assets.

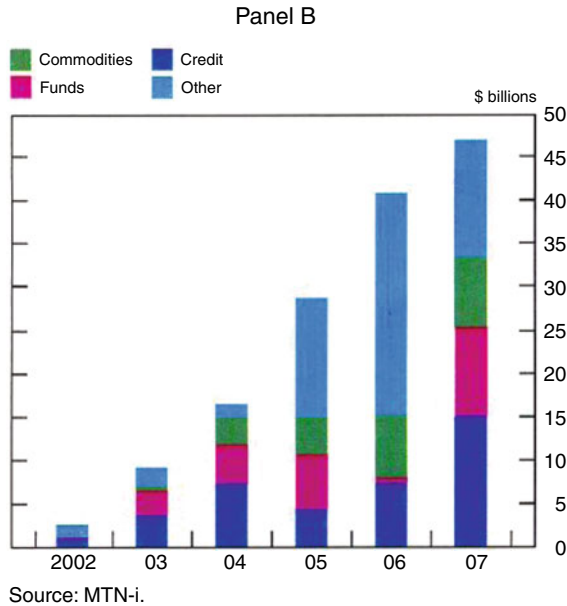
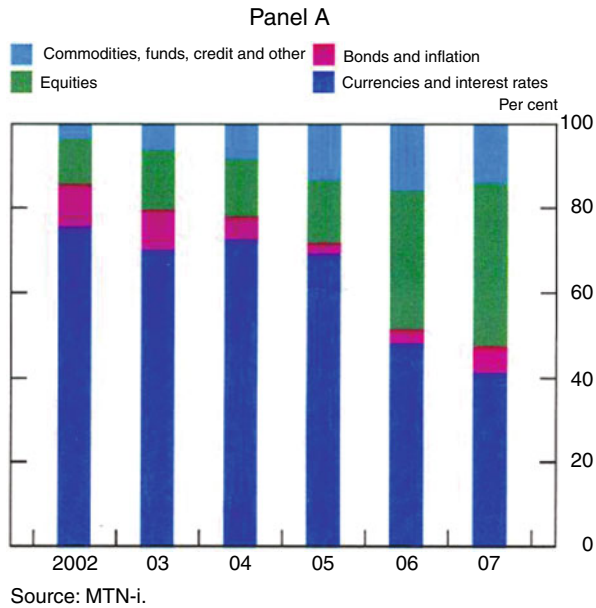
Figure 3 Panel A shows the evolution of shares of structured note issuance by type of risky asset. The share of notes linked to equity and alternative assets (such as commodities, etc.) have increased significantly from 2002 to 2007. In 2007, the shares of currencies and equities are both around 40%, alternative assets occupy more than 10% and bonds have the least share. Figure 3 Panel B further illustrates the evolution of issuance of structured notes linked to alternative assets. The total issuance has grown five times, from around US\$ 5 billion in 2002 to over US\$ 45 billion in 2007.

As for market participants, according to Pain and Rand (2008), portfolio insurance products are more prevalent in Europe than in the United States. The key issuers (sellers) are typically large investment banks that can provide the necessary structuring and marketing expertise. The main investors include both institutional and retail clients, such as pension funds, insurance companies, private banks who purchase products for onward sale to their clients, and high net wealth individuals. Among which, the institutional investors are especially important in continental Europe.

62.3.2 Recent Modifications in CPPI Mechanisms

Since the re-emergence of portfolio insurance strategies, CPPI investments have evolved to incorporate various different features. Of particular importance are the key components in the design of CPPI investments, i.e., the floor, the multiplier, and the exposure to risky assets.

Fig. 3 Structured note issuance by type of risky asset



62.3.2.1 Modifications in Floor

Ratchet floor: The performance of CPPI investments is price dependent, any gains at a particular point in time may still be lost if the underlying asset price subsequently falls. To address this, products with a “ratchet” mechanism have been introduced to

lock in a proportion of upside performance. More specifically, the floor can be set as a percentage of the historical maximum net asset value of the portfolio. That is, the floor jumps up with the portfolio value in order to reduce the risky asset allocation when the market peaks. The time-invariant portfolio protection (TIPP) strategy proposed by Estep and Kritzman (1988) reflects this variation.

Margin floor: The converse to the above case is when the market falls substantially for the cushion to approach zero, the exposure to the risky asset also approaches zero. Under a conventional CPPI, this leads to “cash-out” where the portfolio is fully invested in the risk-free asset, with no possibility of regaining any exposure. One method to counter this is to adjust the starting floor value, by augmenting this by a margin, which can later be used to help regain some exposure after a near cash-out situation. This feature is included in the structured note example (1) below.

Straight-line floor: The floor in a conventional CPPI is a zero-coupon bond, which is sensitive to the level of interest rates. When interest rates fall, the floor would rise and the allocation switches away from the risky asset. This in turn would limit the subsequent potential upside from the CPPI. However, this sensitivity can be removed by allowing the floor to vary linearly with time. The Structured Note Example (3) contains this feature.

62.3.2.2 Modifications in Multiplier

Rather than having a constant multiplier in a conventional CPPI, dynamic portfolio insurance (DPI) allows for the multiplier to vary over time in relation to the volatility of the risky asset and reflects investors’ appetite for risk. In fact, DPI is a cushion management technique that quantifies the level of risk born by the investment in the risky asset. There is often a maximum level for the multiplier, which can be determined by stress testing on the returns of the risky asset using extreme value theory (EVT). The risk-based portfolio insurance strategies in the earlier section reflect this feature.

62.3.2.3 Modifications in the Exposure to Risky Assets

Constraints on the investment level: Another method to avoid the “cash-out” situation when a market falls is to incorporate a minimum level of investment in the risky asset. On the contrary, to avoid unbounded investment in the risky asset as a market rallies, a maximum investment level can also be imposed. This feature can be found in Examples (1), (3), and (4).

Constraints on leverage: Exposure to the risky asset of more than the initial available funds can be achieved by allowing borrowing. There may be constraints on how much can be borrowed. This feature is incorporated in Examples (1) and (4).

Volatility Caps: Some CPPI products include mechanisms that allow the percentage exposure to the risky asset to be reduced if its realized volatility exceeds a certain level. This feature is embedded in Example (2), when realized volatility is greater than the implied volatility, which leads to the value of Reserve Account smaller than 4.5%, the Note stops investing in the Risky Asset and switches into zero coupon bond.

62.3.3 Examples of CPPI Structured Products

Many of the issuances of principal-protected products are designed as medium-term structured notes or deposits, with maturities ranging from 5 to 15 years. The model offers the flexibility to build in additional features, such as zero-coupon, fixed or contingent coupons, and any variations mentioned above. Table 6 lists some real examples of CPPI structured products.

Table 6 Examples of CPPI structured products

Product	(1) Lehman brothers best profile SPA note
Protection	100% at maturity
Currency/tenor	USD/7 years
Participation rate (PR)	50%
Each year end coupon (Cpn_N)	At the end of year 1–6, the Note may pay a coupon with an amount as $Cpn_N = PR * \text{Max}[0\%; \text{Max}(BA1_N, BA2_N, BA3_N) - 100\%]$
	where BAi_N is the value of the i th balanced account at the end of the N th year
	After coupon payment, each of the 3 BAi will be adjusted by deducting the coupon amount for the relevant year
Redemption amount (RA)	At maturity, the Note will redeem with an amount calculated as $RA = \text{Max}[100\%; \text{Max}(BA1_F, BA2_F, BA3_F)]$
	where BAi_F is the value of the i th balanced account at maturity
Balanced account	Three portfolios dynamically allocated between premium assets, safe asset, and leverage borrowing facilities, according to a CPPI mechanism
Premium assets	In respect of each Balance Account, a Basket of fund composed of BA1 (Growth): 50% Equity + 30% Emerging Bond + 20% Global Bond BA2 (Balanced): 1/3 Equity + 1/3 Emerging Bond + 1/3 Global Bond BA3 (Defensive): 20% Equity + 30% Emerging Bond + 50% Global Bond
Underlying	Equity: Merrill Lynch Emerging Europe Fund Emerging Bond: ABN AMRO Global Emerging Bond Fund Global Bond: Templeton Global Bond Fund
CPPI mechanism	
Bond floor	Zero Coupon Bond: Initial Level [75%]
Multiplier	$5\times$
Exposure to risky assets	Initial [125%] ($= 5 * (100 - 75\%)$) Min Exposure [10%], Max Exposure [150%]; Max Leverage [50%]
Product	(2) Lehman brothers AREVO note
Protection	100% at maturity
Currency/tenor	AUD/2.5 years or USD/5.5 years
Participation rate (PR)	100%

(continued)

Table 6 (continued)

Reserve account (RA(t))	At the valuation date of each month (t), the value of Reserve Account is $RA(t) = PR * [Index(t)/Index(t-1) - Strike Price] + LIBOR(t)$ where strike price is 100.1% and the Reserve Account is accumulated to maturity given no cash-out event
Cash-out trigger	4.5% If at the valuation date of any month, the value of Reserve Account is smaller than 4.5%, the Note stops investing in the Index and switches into zero coupon bond
Redemption amount	If there is no cash-out event, the Note will redeem with an amount as $RA = 100\% + \text{Max}[RA(F); 0\%]$ where RA(F) is the cumulated value of the Reserve Account at maturity If cash-out event occurs, the Note will redeem with an amount as $RA = 100\%$
Underlying	S&P500 Volatility Arbitrage Index [Bloomberg: SPARBV Index] The Index records the value of an Arbitrage Relative Volatility (AREVO) strategy, which systemically sells implied volatility and buys realized volatility through Variance Swaps
CPPI mechanism	
Bond floor	Synthetic Zero Coupon Bond: Initial Level [83.5%]
Multiplier	1 ×
Exposure to risky assets	Initial [13.5%] (= 1 * (100–83.5%) – upfront fee 3%)
Product	(3) Credit Suisse principal protected note
Protection	100% at maturity
Currency/tenor	EUR/8 years
Participation rate (PR)	100%
Reference portfolio value (RPV(t))	The Reference Portfolio is defined for calculating the Note Final Redemption Amount and comprises the following assets in proportions as determined by application of the Rebalancing Event (1) The Fund Portfolio (FP) (2) The Fixed Income Portfolio (FIP) On each Valuation Day (t), the Reference Portfolio Value is calculated as If $W(t-1) > 100\%$ $RPV(t) = (1 - \text{Fee} * N(t)/365) * [W(t-1) * FP(t)/FP(t-1) + (1 - W(t-1)) * (1 + (LIBOR(t-1) + \text{Spread}) * N(t)/360)] * RPV(t-1)$ If $W(t-1) \leq 100\%$ $RPV(t) = (1 - \text{Fee} * N(t)/365) * [W(t-1) * FP(t)/FP(t-1) + (1 - W(t-1)) * (FIP(t)/FIP(t-1)) * RPV(t-1)]$
Rebalancing process	A Reference Portfolio Rebalancing Event occurs if, on any Valuation Day (t), either of the following two tests is met $W(t) > (\text{Target Exposure}(t) + \text{Band Del})$ $W(t) < (\text{Target Exposure}(t) - \text{Band Lev})$

(continued)

Table 6 (continued)

	In which case, the weight within the Reference Portfolio will be adjusted on the second Valuation Day following such Valuation Day (t) $W(t + \text{lag}) = \text{Target Exposure}(t); \text{ Otherwise}$ $W(t + \text{lag}) = W(t)$ where $W(0) = W(1) = 100\%$, lag is 2, Band Del and Band Lev are 5%
Target exposure	The target exposure to the Fund Portfolio on each valuation day (t) $TE(t) = \text{Min}\{\text{MaxE}, \text{Max}[\text{MinE}, \text{Multiplier} * (\text{RPV}(t) - \text{BF}(t))/\text{RPV}(t)]\}$ where $\text{MaxE} = 175\%$, $\text{MinE} = 0\%$, $\text{Multiplier} = 4$
Underlying	Fund Portfolio: To be determined [TBD] in EUR Fixed Income Portfolio: The synthetic EUR discounted deposits will redeem on the Final Valuation Date at 100% of its notional amount
CPPI mechanism	
Bond floor	Synthetic Zero Coupon Bond: Initial Level [75%], Final Level [100%] On each month end (t), Bond Floor (BF(t)) is defined by a straight line between the initial and the final valuation dates
Multiplier	$4 \times$
Exposure to risky assets	Initial [100%] ($= 4 * (100 - 75\%)$) Min Exposure [0%], Max Exposure [175%]
Product	(4) Deutsche bank principal protected deposit
Protection	100% at maturity
Currency/tenor	TWD/5 years
Participation rate (PR)	Dynamically adjusted
Balanced account NAV (NAV(t)%)	The initial Principal Amount (NAV(0)%) of TWD 1,000,000 will be dynamically allocated between (1) db X-Harmony II Index (the “Fund Shares”), initial allocation [63%] (2) TWD Zero Coupon Bonds, initial allocation [37%], i.e. $\text{BF}(0)\% = 90.55\%$ (3) FX Swap: rolling 1-month FX swap to buy/sell USD/TWD executed through the onshore swap market Each Month (t), the percentage of Exposure to Fund Shares is equal to $E(t)\% = 6.67 * \text{Distance}(t)\% = 6.67 * (\text{NAV}(t)\% - \text{BF}(t)\%)$ Subject to a minimum exposure = 20% If $E(t)\% = 100\%$, all assets will be invested in Fund Shares If $E(t)\% > 100\%$, borrowing will be made subject to a maximum of 100% of initial Principal Amount, charged at USD LIBOR + 0.65% p.a. If the Distance(t)% ever falls below 5% (a “Trigger Event”), all Fund Shares will be sold and the Balanced Account will only invest into Zero Coupon Bonds until maturity
Rebalancing process	To allow stability of participation, an upper and lower rebalancing band are designed

(continued)

Table 6 (continued)

Redemption amount (RA)	At maturity, the Deposit will redeem with an amount calculated as $RA = \text{Max}[100\%; \text{NAV}(F)\%/\text{NAV}(0)\%] * \text{Outstanding Principal Amount}$
Underlying	db X-Harmony II Index The index tracks 3 funds of hedge funds, expertly managed by leading advisers
CPPI mechanism	
Bond floor	Zero Coupon Bond: Initial Level [37%]
Multiplier	$6.67 \times$
Exposure to risky assets	Initial [63%] Min Exposure [20%], Max Exposure [200%]; Max Leverage [100%]

62.4 Implications for Financial Market Stability

62.4.1 Amplification of Market Price Movements

The portfolio insurance strategies mentioned in this chapter share a common “buy high and sell low” rule for the risky asset. When the underlying asset markets move in one particular direction, either trending up or down, these actions could conceivably have feedback effects in markets that amplify price developments. If the underlying asset markets are deep and liquid, the feedback effects may be limited. However, if the underlying asset markets are less liquid (for example, CPPI products linked to hedge funds on credit derivatives), even small actions may reinforce market prices. Moreover, some underlying assets that appear less correlated in normal conditions may become more correlated in stress situations when CPPI products all look to reduce risky positions simultaneously. Taken together the stress situations together with the inherently less liquid nature of certain underlying asset markets, the feedback impacts would be large.

Duarte (2008) shows that not only the risky asset prices are amplified, the dynamic hedging strategy adopted by MBS portfolio managers is analogous to a CPPI mechanism, which also intensifies the price movements of riskless assets such as Treasury Notes. Suppose, for example, that an MBS portfolio manager hedges the portfolio duration risk with a short position in Treasury Notes. If interest rates drop, the MBS portfolio duration decreases due to a higher probability of refinancing. As a result, the manager will have a portfolio with negative duration. To adjust portfolio duration back to zero, the manager must buy Treasury Notes. If, on the other hand, interest rates increase, the MBS portfolio duration increases and the manager must short additional Treasury Notes. In other words, the dynamic hedging strategy for MBS portfolio makes the manager buy Treasury Notes when prices are going up and sell Treasury Notes when prices are going down. This hedging behavior has the effect of reinforcing both the movement of Treasury Notes prices and their volatility. Table 7 shows some statistics on mortgage-related and Treasury securities. Since

Table 7 Statistics on mortgage-related and treasury securities

	1995	1999	2000	2001	2002	2003
Outstanding residential mortgage-related security						
Agency MBSs	1,554	2,252	2,439	2,770	3,089	3,364
Nonagency MBSs	224	455	500	591	692	843
Total	1,779	2,707	2,939	3,362	3,781	4,207
Outstanding US treasury debt						
Bills	761	737	647	811	889	929
Notes/bonds	2,531	2,429	2,184	2,017	2,170	2,470
Total	3,292	3,166	2,831	2,828	3,059	3,399

year 2000, the outstanding amount of residential mortgage-related security becomes larger than that of Treasury Debt. Provided that bond prices are affected by flows in the Treasury market, the huge MBS hedging flows will have a big impact on the riskless bond prices.

Fortunately, according to Pain and Rand (2008), market contacts reveal that the amount of portfolio insurance-related dynamic hedging flows still remain modest relative to the size of the corresponding underlying asset markets, even in those new asset classes. Moreover, market contacts do not perceive portfolio insurance as a significant factor during the 2007 financial market turmoil as that perceived as the driving force during the 1987 stock market crash.

62.4.2 Gap Risk

The biggest potential risk of implementing a portfolio insurance strategy is that the payoffs are particularly vulnerable to sudden huge drops in the risky asset prices before the portfolio can be rebalanced. In such cases, the value of the insured portfolio would fall below the floor; however, a hard guarantee, such as 100% principal protection, to the investors must be fulfilled at the end of investment horizon. Therefore, the issuers of CPPI products would suffer a gap risk. In particular, if the issuers issue several CPPI products written on different underlying assets, and those seemingly uncorrelated assets suddenly become much more correlated in stressed conditions (such as the credit crunch led by subprime mortgage crisis since September 2007), then the scale of the gap risk may be very much underestimated. The unexpected huge losses may jeopardize the issuers and the stability of financial markets.

Figure 4 shows two of the simulated outcomes of Example (3). Assume that the price of the underlying risky asset follows a geometric Brownian Motion with a drift of 7% (p.a.) and a volatility term. When the volatility increases from 0% to 30%, the reference portfolio value would fall below the bond floor; however, a 100% principal protection to the investors must be fulfilled at the end of investment horizon. Therefore, the issuers would suffer a gap risk.

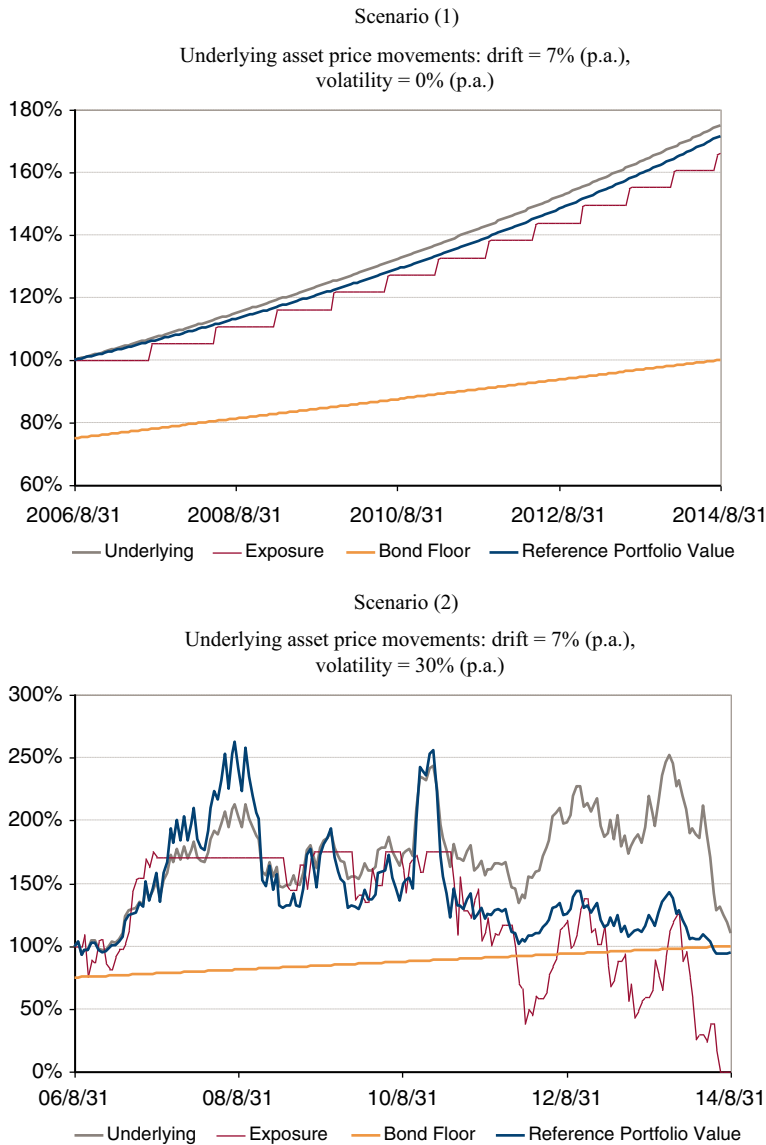


Fig. 4 Simulated Performances of Structured Note Example (3)

In theory, the issuer of CPPI products can hedge such exposure to gap risk by options. The issuer needs to model the likely worst-case move in the risky asset price before the next rebalancing opportunity and build the cost of this implicit option into the premiums and fees charged to the investors or provided by capital. But, in practice, the pricing of such options can be quite complex, because the issuer does not exactly know the underlying asset price processes and their correlations in

stressed markets. Furthermore, it is difficult for the issuers to find available options through which to hedge their exposures.

Pain and Rand (2008) report that some issuers of CPPI products create securities that package up the gap risk and sell these to investors, including private banks and funds. But these structured derivatives are not popular due to the complicated nature of risk. Other issuers seek to limit their exposure to gap risk by other imprecise hedging vehicles or so-called proxy hedging. But, this in turn exposes the issuers to the basis risk.

62.5 Empirical Comparison of Alternative Portfolio Insurance Strategies

In the late 1980s and early 1990s, much of the research in this area focused on simulated comparisons of the performances among alternative portfolio insurance strategies. In the 2000s, most studies use empirical data to analyze the circumstances where one strategy would outperform the others based on some modifications on certain portfolio insurance mechanisms, such as volatility input, rebalancing intervals, etc. The performances are typically evaluated by floor protection – whether they achieve the desired floor at the end of the investment horizon, the cost of insurance – the opportunity cost of forfeiting upward markets, and the Sharpe ratio. Listed below are some of the studies that have appeared in academic journals.

62.5.1 Zhu and Kavee (1988)

They evaluate and compare the performances of the two traditional portfolio strategies, the synthetic put approach of Rubinstein and Leland (1981) and the constant proportion approach of Black and Jones (1987). They employ a Monte Carlo simulation methodology, assuming log normally distributed daily returns with an annual mean return of 15% and paired with different values for market volatility, in order to discover whether these strategies can really guarantee the floor return and how much investors have to pay for the protection.

Both strategies are able to reshape the return distribution so as to reduce downside risk and retain a certain part of the upside gains. However, they demonstrate that a certain degree of protection can be achieved at a considerable cost. There are two types of costs in implementing a portfolio insurance strategy. The first is the explicit cost, that is, the transactions costs. The other is the implicit cost, which is the average return forgone in exchange for protection against the downside risk. When the market becomes volatile, the protection error of the synthetic put approach increases, and the transaction costs may be unbearable. On the contrary, while the constant proportion approach may have lower transactions costs, its implicit cost may still be substantial.

62.5.2 Perold and Sharpe (1988)

Using simulated stocks and bills prices, they examine and compare how the four dynamic asset allocation strategies, namely, the buy-and-hold, constant mix, CPPI and OBPI, perform in bull, bear and flat markets, and in volatile and not-so-volatile markets.

CPPI and OBPI strategies sell stocks as the market falls and buy stocks as the market rises. This dynamic allocation rule represents the purchase of portfolio insurance and has a convex payoff function, which results in a better downside protection and a better upside potential than a buy-and-hold strategy. However, they do worse in relatively trendless, volatile markets.

On the contrary, a constant mix strategy – holding a constant fraction of wealth in stocks – buys stocks as the market falls and sells them as it rises. This rebalancing rule effectively represents the sale of portfolio insurance and has a concave payoff function, which leads to less downside protection than, and not as much upside as, a buy-and-hold strategy. However, it does best in relatively trendless and volatile markets.

They suggest that no one particular type of dynamic strategy is best in all situations. Financial analysts can help investors understand the implications of various strategies, but they cannot, and should not, choose a strategy without a substantial understanding of an investor's circumstances and desires.

62.5.3 Rendleman and O'Brien (1990)

They address the issue that the misestimation of volatility input can have a significant impact on the final payoffs of a portfolio using a synthetic put strategy. In an OBPI strategy, the daily portfolio adjustments depend on the delta of the put option on the risky asset. In the original Black and Scholes (1973) valuation equation, the delta at each day is a function of the following variables:

$$D_t = f(S_t, \bar{k}_0, \bar{r}_0, \sigma_0, T) \quad (16)$$

where S_t is the price of the risky asset at date t , k_0 the strike price, r_0 the annual continuously compounded riskless rate of interest, σ_0 the ex-ante volatility parameter at the beginning of the insurance period, and T the maturity of the option.

Assume that at the beginning of the insurance period, one manager predicts a high-volatile market and a 20% annualized volatility, while a second manager believes a low-volatile market and a 10% volatility. The delta for the first manager is higher which means he would buy more insurance and allocate less to a position in the risky asset. Assume that the ex-post volatility turns out to be 10%. The second manager will have made the proper allocation between risky and riskless assets. In contrast, the high-estimate manager will have invested too little in the risky asset. This misallocation would lose the opportunity of participating in the price appreciation in a strong market. Thus, a manager who underestimates volatility will

typically end up buying less insurance than is necessary to ensure a given return, while a manager who overestimates volatility will buy more insurance than is necessary and forgo gains.

As for the issue of portfolio rebalancing, they examine adjustment frequencies by time intervals, namely, daily, weekly, monthly, and bimonthly. The effect (error) is measured as the difference of the horizon insurance values between noncontinuous and continuous trading. They suggest that weekly rebalancing produces an amount of error that appears to be tolerable by most portfolio managers.

More importantly, they address the biggest potential risk of implementing portfolio insurance strategies – the gap risk, by simulating the performance of the OBPI strategy over the period of the October 1987 market crash. They indicate that most insured portfolios would have fallen short of their promised values because the managers would not be able to adjust the portfolio in time before a big drop in the market. The gap risk will be discussed further in the financial stability section.

62.5.4 Loria, Pham, and Sim (1991)

They simulate the performance of a synthetic put strategy using futures contracts based on the Australian All Ordinaries Index for the period from April 1984 to March 1989. Their study contains 20 consecutive nonoverlapping 3-month insurance periods whose expiration dates coincide with the expiration dates of each SPI futures contract traded on the Sydney Futures Exchange. Four implementation scenarios are examined: a zero floor versus 5% floor which correspond to a portfolio return of 0% and –5% per annual, respectively; and a realized volatility versus Leland's (1985) modified volatility. They report that there is no perfect guarantee of loss prevention under any scenario. Even in the scenario with a 5% floor and modified volatility, 2 out of 20 contracts do not meet the desired floor. In addition, the OBPI strategy is most effective under severe market under severe market conditions. In other periods characterized by insignificant market declines, the value of the insured portfolio is below that of the market portfolio. They suggest futures mispricing may be one potential culprit for this outcome.

62.5.5 Do and Faff (2004)

This empirical paper is an extension of Loria et al. (1991). They examine two approaches (the OBPI and CPPI) and conduct simulations across two implementation strategies (via Australia stock index and bills, and via SPI futures and stock index). Furthermore, they consider the use of implied volatility as an input into the model, and the use of a zero floor versus 5% floor. The dataset consists of 59 non-overlapping 3-month insurance periods, which span from October 1987 to December 2002. Thus, their key contributions relative to Loria et al. (1991) are the examination of a futures-based CPPI, a fine-tuned algorithm that allows for

dividend payments, the consideration of ex ante volatility information, and more up-to-date data.

In terms of floor protection, the futures-based portfolio insurance implementation generally dominates its index-and-bill rival in both floor specifications, which reflects the low transaction costs in the futures market. Furthermore, the perfect floor protection is possible when implied volatility is used rather than using ex-post volatilities. From the cost of insurance perspective, the futures-based strategy generally induces a lower cost of insurance than its index-and-bill rival in both floor specifications, which reflects the same reason as above. However, the cost of insurance is higher when implied volatility is used compared to ex-post volatilities. The possible explanation is that the implied volatility is often higher than the same period ex-post volatility, which results in over-hedging. As for the performances between the OBPI and CPPI approaches, there is no strong evidence to distinguish between them. Within the futures-based implementation, the synthetic put appears to dominate the CPPI with respect to floor protection, while the latter appears to slightly outperform in terms of upside participation.

They also examine whether portfolio insurance strategies work under stress conditions by assessing the strategies' effectiveness during tranquil and turbulent periods. Tranquil periods are defined as ones that return more than -4% and have a realized volatility of less than 15% ; violation of either one or both of these conditions is regarded as indicating a turbulent period. All portfolio insurance strategies achieve 100% floor protection during tranquil periods whereas the futures-based OBPI approach records the highest portfolio return. During turbulent times, futures-based portfolio insurance continues to perform quite well. The 1987 stock market crash makes the assessment difficult because of trading halts. However, assuming the futures continued to trade during that crisis, from the algorithm's perspective, the futures-based CPPI maintains a positive return, while the OBPI results in a negative return.

62.5.6 Cesari and Cremonini (2003)

This study is an extensive comparison of a wide variety of traditional portfolio insurance strategies. There are basically five dynamic asset allocation strategies: (1) buy-and-hold (BH), (2) constant mix (CM), (3) constant proportion (without and with the lock-in of profits, CP and CPL), (4) the option-based approach (with three variations, BCDT, NL, PS), and (5) technical strategy (with two kinds of stop-loss mechanism, MA and MA2), therefore, nine strategies in total are considered.

For each strategy, eight measures for risk, return and risk-adjusted performance are calculated, namely, mean return, standard deviation, asymmetry, kurtosis, downside deviation, Sharpe ratio, Sortino ratio and return at risk. The strategies are then compared in different market situations (bear, no-trend, bull markets) and with different market volatility periods (low, medium, and high periods), taking into account transaction costs and discrete rebalancing of portfolios. The three market situations are defined accordingly if market average returns fall into the three ranges:

(−30%, −5%), (−5%, +5%), (+5%, +30%), and the three ranges for the volatile periods are: (10%, 15%), (15%, 25%), and (25%, 30%). Transaction costs are treated in two ways: a proportional cost to the value traded and a correction to the Leland's (1985) option volatility. Two main rebalancing disciplines are used: a time discipline with weekly adjustment, and a price discipline with adjustment only when prices are increased/decreased by 2.5% with respect to the previous rebalance time.

Monte Carlo simulations of MSCI World, North America, Europe, and Pacific stock market returns show that no strategy is dominant in all market situations. However, in bear and no-trend markets, CP and CPL strategies appear to be the best choice. In a bull market or in no-trend but with high volatility market, the CM strategy is preferable. If the market phase is unknown, CP, CPL, and BCDT strategies are recommended. In addition, these results are independent of the volatility level and the risk-adjusted performance measure adopted.

62.5.7 Herold, Maurer, and Purschaker (2005)

By constructing a fixed income portfolio, in which the risky asset is the JPMorgan government bond index and the riskless asset is cash (1-month yield), they compare the hedging performances between the traditional CPPI strategy and the risk-based (specifically, VaR-based) strategy with a 1-year investment horizon that begins at each year from 1987 to 2003. CPPI avoids losses in the bear years of 1994 and 1999. The mean return is inferior to (about 40 bp below) that of the risk-based strategy. CPPI also produces a higher turnover.

62.5.8 Hamidi, Jurczenko, and Maillet (2009)

Although they propose a conditional CPPI multiplier determined either by VaR or by ES, only a VaR-based measure is studied in this empirical work, and the ES measure is absent for future research.

The data set contains 29 years of daily returns of the Dow Jones Index, from 2 January 1987 to 20 May 2005, 4641 observations in total. They use a rolling window of 3,033 returns to estimate the VaR, and there are 1,608 VaRs in the out-of-sample period. They resort to eight methods of VaR calculation: one nonparametric method using the historical simulation approach; three parametric methods based on distributional assumptions, namely, the normal VaR, the RiskMetrics VaR based on the normal distribution, and the GARCH VaR based on the Student-t distribution; four semi-parametric methods using quantile regression to estimate the conditional autoregressive VaR (CAViaR), namely, the symmetric Absolute Value CAViaR, the Asymmetric Slope CAViaR, the IGARCH(1,1) CAViaR, and the Adaptive CAViaR.

According to the 1,608 back-testing results, the Asymmetric Slope CAViaR is the best model to fit the data. After having calculated the VaR values, the conditional multipliers can be determined. The estimations spread between 1.5

and 6, which are compatible with multiple values used by practitioners in the market (between 3 and 8).

Using the time-varying CPPI multipliers estimated by different methods, and a “multi-start” analysis – the fixed 1-year investment horizon beginning at every day of the out-of-sample period, they find that the final returns of these insured portfolios are not significantly different.

62.5.9 Ho, Cadle, and Theobald (2011)

This empirical study presents a complete structure of comparing traditional portfolio insurance strategies (OBPI, CPPI) with modern risk-based portfolio insurance strategies (VaR-, ES-based RBPI). By constructing a currency portfolio, in which the risky asset is the Australian dollar and the riskless asset is a US dollar overnight deposit, they compare the dynamic hedging performances between the traditional and the modern strategies with a 1-year investment horizon that begins at each year from 2001 to 2007.

When implementing the OBPI strategy, the delta is calculated based on the modified Black-Scholes formula, $D_t = f(S_t, \bar{k}_0, r_t, \sigma_t, T)$. That is, daily annualized historical volatilities are used as inputs in the put-replication process instead of the constant ex-ante or the implied volatility in the original model to mitigate the volatility misestimation problem. Besides, the latter two parameters would make the portfolio daily returns more volatile. The interest rates are also updated daily for the same reason. When CPPI is implemented, the possible upper bounds of the multiplier are examined via the extreme value theory, which ranges from 4 to 6 corresponding to different confidence levels. When risk-based approaches are employed, both a historical distribution and a normal distribution with exponentially weighted volatility of risky asset returns are assumed. A daily rebalancing principle is adopted in their research without any modification to show the original results of hedging.

The performances are evaluated from six differing perspectives. In terms of the Sharpe ratio and the volatility of portfolio returns, the CPPI is the best performer, while the VaR based upon the normal distribution is the worst. From the perspective that the return distribution of the hedged portfolio is shifted to the right and in terms of both the average and the cumulative portfolio returns across years, the ES-based strategy using the historical distribution ranks first. Moreover, the ES-based strategy results in a lower turnover within the investment horizon, thereby saving transaction costs.

62.6 Conclusion

A portfolio insurance strategy is a dynamic hedging process that provides the investor with the potential to limit downside risk while allowing participation on the upside so as to maximize the terminal value of a portfolio over a given

investment horizon. This entry looks at the basic concept of portfolio insurance, the alternative dynamic hedging strategies available to the portfolio manager, the recent market developments on the principal-protected structured products, the implications of portfolio insurance on financial market stability, and finally, the empirical comparison of the performances of various portfolio insurance strategies.

References

- Bertrand, P., and J. L. Prigent. 2001. *Portfolio insurance: The extreme value approach to the CPPI method*. Thema University of Cergy, Working paper 2001-A13.
- Bertrand, P., and J.L. Prigent. 2005. Portfolio insurance strategies: OBPI versus CPPI. *Finance* 26 (1): 5–32.
- Black, F., and R. Jones. 1987. Simplifying portfolio insurance. *Journal of Portfolio Management*: 48–51.
- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–659.
- Cesari, R., and D. Cremonini. 2003. Benchmarking, portfolio insurance and technical analysis: A Monte Carlo comparison of dynamic strategies of asset allocation. *Journal of Economic Dynamics & Control* 27: 987–1011.
- Do, B.H. 2002. Relative performance of dynamic portfolio insurance strategies: Australian evidence. *Accounting and Finance* 42: 279–296.
- Do, B.H., and R.W. Faff. 2004. Do futures-based strategies enhance dynamic portfolio insurance? *The Journal of Futures Markets* 24 (6): 591–608.
- Duarte, J. 2008. The causal effect of mortgage refinancing on interest rate volatility: Empirical evidence and theoretical implications. *The Review of Financial Studies* 21 (4): 1689–1731.
- Estep, T., and M. Kritzman. 1988. TIPP: Insurance without complexity. *Journal of Portfolio Management* 14 (4): 38–42.
- Garman, M.B., and S.W. Kohlhagen. 1983. Foreign currency option values. *Journal of International Money and Finance* 2: 231–237.
- Hakanoglu, E., R. Kopprasch, and E. Roman. 1989. Constant proportion portfolio insurance for fixed-income investment: A useful variation on CPPI. *Journal of Portfolio Management* 15 (4): 58–66.
- Hamidi B., E. Jurczenko, and B. Maillet. 2009. A CAViaR modelling for a simple time-varying proportion portfolio insurance strategy. *Bankers, Markets and Investors*, September–October, 4–21.
- Herold, U., R. Maurer, and N. Porschaker. 2005. Total return fixed-income portfolio management: A risk-based dynamic strategy. *Journal of Portfolio Management*: 31 (3): 32–43.
- Ho, L.C., J. Cadle, and M. Theobald. 2011. An analysis of risk-based asset allocation and portfolio insurance strategies. *Review of Quantitative Finance and Accounting* 36 (2): 247–267.
- Leland, H. 1985. Option pricing and replication with transaction costs. *Journal of Finance* 40: 1283–1301.
- Loria, S., T.M. Pham, and A.B. Sim. 1991. The performance of a stock index futures-based portfolio insurance scheme: Australian evidence. *Review of Futures Markets* 10 (3): 438–457.
- Pain, D., and J. Rand. 2008. Recent developments in portfolio insurance. *Bank of England Quarterly Bulletin* Q1: 37–46.
- Perold, A. R. 1986. *Constant proportion portfolio insurance*. Harvard Business School, Working paper.
- Perold, A.R., and W.F. Sharpe. 1988. Dynamic strategies for asset allocation. *Financial Analysts Journal* 44 (1): 16–27.

- Rendleman, R.J., Jr., and T.J. O'Brien. 1990. The effects of volatility misestimation on option-replication portfolio insurance. *Financial Analysts Journal* 46 (3): 61–70.
- Rubinstein, M., and H.E. Leland. 1981. Replicating options with positions in stock and cash. *Financial Analysts Journal* 37: 63–72.
- Zhao, Y., and W.T. Ziemba. 2000. A dynamic asset allocation model with downside risk control. *Journal of Risk* 3 (1): 91–113.
- Zhu, Y., and R.C. Kavee. 1988. Performance of portfolio insurance strategies. *Journal of Portfolio Management* 14 (3): 48–54.



Time-Series and Cross-Sectional Tests of Asset Pricing Models

63

Kyung-Jin Choi, Dongcheol Kim, and Soon-Ho Kim

Abstract

This chapter reviews the methodologies of testing asset pricing models which are dominantly used in the literature: time-series regression tests and cross-sectional regression tests. We provide some explanations for the test procedure of time-series regression tests and cross-sectional regression tests. We discuss individual t -test, the joint F -test by Gibbons, Ross, and Shanken (*Econometrica* 57: 1121–1152, 1989), and tests based on the generalized method of moments estimation. We also explain the two-pass test methodology and discuss the errors-in-variables problem which occurs inevitably in the two-pass methodology.

Keywords

Asset pricing models · Cross-sectional tests · Time-series tests

K.-J. Choi

Korea University Business School, Seoul, South Korea

D. Kim (✉)

Korea University Business School & SolBridge International School of Business at Woosong University, Seoul, South Korea

e-mail: kimdc@korea.ac.kr

S.-H. Kim

Pukyong National University, Busan, South Korea

e-mail: kimsoonho@pknu.ac.kr

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_63

1467

63.1 Introduction

It is very important to determine an appropriate asset pricing model since it is essential in estimating cost of capital for firms and discount rates for project cash flows and evaluating the performance of managed portfolios, etc. Since Sharpe (1964) proposed capital asset pricing model (CAPM), many theoretical and empirical asset pricing models have been suggested in the literature. In particular, numerous empirical asset pricing models have been suggested with some arguments. It is important, therefore, to determine an asset pricing model among these that describes well the intertemporal and cross-sectional dynamics of asset returns. To do so, it is essential to apply a correct test procedure.

This chapter reviews the methodologies of testing asset pricing models (see also Kim 2011). We provide some explanations for two testing methodologies dominantly used in the literature: time-series regression tests and cross-sectional regression tests. Time-series tests focus on the intertemporal explanatory power of returns on given factors for returns on test assets returns and on the pricing error of the given asset pricing model. Meanwhile, the cross-sectional tests focus on the explanatory power of factor loadings for asset returns in the cross-section. We explain individual t -test and the joint F -test by Gibbons et al. (1989) which are used in time-series tests. We also explain the two-pass test methodology that is used in cross-sectional test, and discuss the errors-in-variables problem which occurs inevitably in the two-pass methodology. Moreover, we provide some explanations of tests based on the generalized method of moments estimation and tests based on the Hansen and Jagannathan (1997) distance.

This chapter is organized as follows. Section 63.2 explains time-series regression tests, Sect. 63.3 discusses tests based on the generalized method of moments estimation, and Sect. 63.4 explains cross-sectional regression tests and discusses the errors-in-variables problem. Section 63.5 concludes.

63.2 Time-Series Tests

63.2.1 The Factor Model

Consider the following return generating process of the N assets. The returns of an asset i are generated by the following linear equation:

$$R_{it} - R_{ft} = \alpha_i + \beta_i' F_t + \varepsilon_{it}, \quad (1)$$

where R_{it} is the return on the asset i ($i = 1, \dots, N$) at time t ($t = 1, \dots, T$), R_{ft} is the risk-free return, ε_{it} is the residual or idiosyncratic return with mean zero and covariance matrix Σ_ε , β_i is a $(K \times 1)$ vector of the factor loadings, and F_t is a $(K \times 1)$ vector

of the factor portfolio returns. If the model (1) is well specified, the equilibrium pricing equation is described as

$$E(r_{i,t}) = \gamma' \beta_i \quad (2)$$

where γ is a $(K \times 1)$ risk premia vector, (the j -th element γ_j is the risk premium of the j -th risk factor), and $r_{i,t}$ is the return of asset i in excess of the riskless rate of return ($=R_{it} - R_{ft}$). The above pricing equation of (2) has one restriction that the intercepts should be zero. The intercept from the time-series factor model of Eq. 1 is interpreted as the pricing error. It is also dubbed as the Jensen alpha. Thus, if $\alpha_i = 0$ for all test assets, we interpret this as evidence that supports the validity of a given factor model or the mean-variance efficiency of a given portfolio (or a linear combination of given factor portfolios).

63.2.2 Individual t -Test

The typical examples of time-series tests using Eq. 1 are Black et al. (1972) (hereafter BJS), and Fama and French (1993). BJS estimate the time-series regression model of Eq. 1 for each of test assets and examine whether the intercept estimate α_i is significantly different from zero to test the validity of the CAPM. They use ten beta-sorted portfolios as test assets. Fama and French (1993) also estimate Eq. 1 with three stock factor portfolios and two bond market portfolios by using 25 size-BM sorted portfolios as test assets. These two studies and many other studies use individual t -statistic of the intercept estimates to test the validity of a given factor model.

63.2.3 Joint F -Test

To test whether the intercept estimates of the time-series regression (1) are jointly different from zero, Gibbons et al. (1989) (hereafter GRS) develop an F -tests. If the given asset pricing model or factor model is well specified, the intercept estimates should not be different from zero. Therefore GRS F -statistic is to test a joint hypothesis $H_0 : \alpha = (\alpha_1, \alpha_2, \dots, \alpha_N) = 0$. Since the null hypothesis indicates that some linear combination of the factor portfolios is on the mean-variance efficiency boundary, the GRS test is in fact to test whether the linear combination of the factor portfolios is on the mean-variance efficient boundary.

Assuming the error terms are jointly normally distributed with mean zero and nonsingular error covariance matrix Σ_e , GRS suggest the following test statistic for the null hypothesis:

$$W = \left[\frac{T(T-N-K)}{N(T-K-1)} \right] \left[\frac{\hat{\alpha}' \widehat{\sum}_e^{-1} \hat{\alpha}}{1 + \hat{\vartheta}^2} \right] \quad (3)$$

$$\sim F(N, T-N-K),$$

where $\widehat{\sum}_e$ is the unbiased residual covariance matrix, $\hat{\vartheta}^2 = \bar{F}' \widehat{\sum}_F^{-1} \bar{F}$, $\bar{F}$ is a $(K \times 1)$ vector of average returns of the factor portfolios $\left(= (1/T) \sum_{t=1}^T F_t \right)$, $\widehat{\sum}_F$ is the covariance matrix of the factor portfolio returns. Here, $\hat{\vartheta}$ is the ex post maximum Sharpe ratio of the K factor portfolios. The GRS test statistic of Eq. 3 has a nice geometric interpretation. Since $\hat{\alpha}' \widehat{\sum}_e^{-1} \hat{\alpha} = \hat{\vartheta}^{*2} - \hat{\vartheta}^2$ where $\hat{\vartheta}^*$ is the ex post maximum Sharpe ratio of the N test assets and the K factor portfolios, the GRS test statistic equals

$$W = \left[\frac{T(T-N-K)}{N(T-K-1)} \right] \left[\left(\frac{1 + \hat{\vartheta}^{*2}}{1 + \hat{\vartheta}^2} \right) - 1 \right] \quad (3a)$$

$$\sim F(N, T-N-K).$$

Thus the GRS test determines whether $|\hat{\vartheta}^{*2}|$ is statistically greater than $|\hat{\vartheta}^2|$. As the squared Sharpe ratio of the tangency portfolio constructed from factor portfolios $\hat{\vartheta}^2$ decreases, the GRS test statistic increases, which is evidence against the mean-variance efficiency of some linear combination of the factor portfolios. If the given factor model is the one-factor model with the market index portfolio such as Standard and Poors 500 Index, the GRS test is to test the mean-variance efficiency of the given market index portfolio.

63.3 Tests Based on Generalized Methods of Moments (GMM)

63.3.1 Introduction of the GMM

The time-series tests based on individual t -tests and the GRS tests assume that asset returns or the error terms of Eq. 1 are normally distributed. It is well known, however, that asset returns are not normally distributed. In this circumstance, the generalized method of moments (GMM) developed by Lars Hansen (1982) is useful. The GMM refers to the estimation which uses the sample moment conditions from the population moment conditions. The main advantage of the GMM is that it makes empirical tests of an economic theory possible under a general framework. Specifically, the GMM makes nonlinear asset pricing models be empirically testable without linearization. The GMM is computationally convenient without strong distributional assumptions, and thus it can be used under serial correlation and conditional heteroskedasticity in making inferences. Moreover, it does not have to

assume data generating processes. These advantages are the reason that the GMM is one of the most frequently applied methodologies in finance.

It should be mentioned, however, that the GMM estimators are only asymptotically valid. Ferson and Foerster (1994) find that a two-stage GMM approach tends to reject the models too often in a moderate sample size. Arellano and Bond (1991) also find that the estimated standard errors of the GMM estimators are severely downward biased in a small sample size. Ahn and Gadarowski (2004) show that asset pricing tests based on the GMM estimation reject the correct model too often. In other words, like other tests based on asymptotic estimators, tests based on the GMM estimation have too large Type I error.

63.3.2 Overview of the GMM Estimation

An economic model can be specified as a vector of population moment conditions: $E[g(x_t, \theta)] = E[g_t(\theta)] = 0$, where x_t is a vector of P variables for $t = 1, \dots, T$, and θ is a parameter vector of K unknown parameters, $g(\cdot)$ is a vector of N functions, with the necessary order condition for identification such that $K \leq N$. Since the order condition is a necessary condition for identification, a failure of the order condition means that the model is not identified. When the model is overidentified, Hansen's (1982) J test can be implemented as using the overidentifying restrictions. The exactly identified model cannot be tested whether the model is correctly specified or not.

Let $g_T(\theta) = \frac{1}{T} \sum_{t=1}^T g_t(\theta)$ be the corresponding sample moments to the population moments. In the asset pricing test framework, $g_T(\theta)$ can be interpreted as a vector of pricing errors. Then, the GMM estimator $\hat{\theta}$ for θ minimizes the following quadratic function with respect to θ .

$$Q = [g_T(\theta)]' W [g_T(\theta)], \quad (4)$$

where W is a weighting matrix which is the positive definite. The weighting matrix W implies the weight to be given for each moment. If W is an identity matrix, the GMM treats all moments symmetrically. This means that all assets have the same degree of importance in estimating the parameters and that the objective function Q is simply to minimize the sum of squared pricing errors.

Hansen (1982) proposes the optimal weighting matrix as follows. The optimal weighting matrix produces the estimates with the smallest asymptotic variance.

When $E[g_t(\theta)] = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T g_t(\theta) = p_{T \rightarrow \infty}^{lim} g_T(\theta)$ and the central limit theorem is applied,

$$\sqrt{T} g_T(\theta) \xrightarrow{d} N(0, V) \text{ as } T \rightarrow \infty, \quad (5)$$

where $V = \sum_{j=-\infty}^{\infty} E[g(x_t, \theta) g(x_{t+j}, \theta)']$ which is a nonsingular covariance matrix. V is the spectral density matrix of $g_T(\theta)$. When $W = V^{-1}$, Hansen (1982) finds that

the GMM estimates are efficient. A lower bound for the asymptotic variance of the GMM estimates can be achieved as choosing $\mathbf{W} = \mathbf{V}^{-1}$. Hansen (1982) shows that the asymptotic distribution of the GMM estimator is given by

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1}) \text{ as } T \rightarrow \infty, \quad (6)$$

$$\text{where } D = E\left[\frac{\partial g_t(\theta)}{\partial \theta}\right] = \lim_{T \rightarrow \infty} \frac{\partial g_T(\theta)}{\partial \theta}.$$

To calculate the efficient GMM estimate, the consistent estimator $\hat{\mathbf{V}}$ is needed. Newey and West (1987) and Cochrane (2004) show the estimation of $\mathbf{V}$ in the form of spectral density matrix. A two-stage procedure can be employed to obtain the efficient GMM estimator. In the first stage, a consistent estimator $\hat{\mathbf{V}}$ is obtained by using the identity matrix as the weighting matrix. In the second stage, the efficient GMM estimator can be produced using $\hat{\mathbf{V}}^{-1}$ as the weighting matrix. Cochrane (2004) and Hayashi (2000) describe in details. Tests for the model misspecification can be implemented as examining the extent of the pricing errors $g_T(\hat{\theta})$. Hansen (1982) proposes J -statistics as follows:

$$J = T g_T(\hat{\theta})' \mathbf{V}^{-1} g_T(\hat{\theta}) \text{ and } J \xrightarrow{d} \chi^2(N - K) \text{ as } T \rightarrow \infty. \quad (7)$$

If the model is exactly identified, then $J = 0$.

Generally, the efficient GMM estimates are known to have poorer small sample properties. The model misspecification test based on the efficient GMM estimates also rejects the null hypothesis too often. Therefore, sufficiently large number of time-series return observations (T) relative to the number of assets (N) is required to obtain the consistent estimators. As the number of assets becomes large relatively to T , it is difficult to estimate the precise covariance matrix of returns. To avoid this problem, one approach is to group the assets into a small number of portfolios. Ferson and Foerster (1994) show that a GMM approach still tends to reject the models too often in moderate sample size. Other solutions are also proposed to improve the inference problem caused by the underestimation of the standard error of the GMM estimator. Ren and Shimotsu (2009) propose an improved method to compute the weighting matrix by applying the shrinkage method (Ledoit and Wolf (2003)). Windmeijer (2005) suggests the correction term to generate the more precise estimator of the standard error. Windmeijer (2005) adjusts the bias from the extra variation in small samples and reports the improved results.

63.3.3 Moments Conditions for the Asset Pricing Tests

Once the moment conditions are identified, the estimation and the inference are straightforward. Generally, the optimization is performed by the numerical method.

The closed-form solutions for the parameters are obtained only for some special cases in the GMM method.

The GMM method can be used for both the beta representation and the stochastic discount factor (SDF) representation in the asset pricing model. The asset pricing model under the beta representation is as follows:

$$E[r_t] = \gamma_1\beta \text{ or } E[R_t] = \gamma_0 + \gamma_1\beta, \quad (8)$$

where r_t is the excess return, R_t is the raw return, γ_0 is the return of the zero beta asset, γ_1 is the factor risk premium, and $\beta = \text{Cov}(R_t, f_t) / \text{Var}(f_t)$ is the sensitivity of the asset return to the factor. The SDF representation is given by

$$E[r_t m_t] = 0 \text{ or } E[R_t m_t] = 1, \quad (9)$$

where m_t is referred to as a stochastic discount factor. For example, the stochastic discount factor is given by $m_t = \lambda_0 + \lambda_1 f_t$ for a one-factor linear pricing model. If the model is true, then $m_t \in M_t$, where M_t is the set of the true stochastic discount factors.

The SDF representation using the GMM becomes popular in testing asset pricing models, since it can be used within the general framework to analyze either linear or nonlinear pricing models. However, there are some concerns that the generality of the SDF method using the GMM comes at the cost of efficiency in two aspects: estimation of the parameters and power in the specification tests. Kan and Zhou (1999) point out that the SDF method using the GMM has two problems when asset returns are generated by a linear factor model. The first is that the risk premium estimate from the SDF method is not reliable. The second problem is that the model specification test using the SDF method has too low power in detecting the misspecified model. But Jagannathan et al. (2002) argue that the SDF method using the GMM is asymptotically as efficient as the beta method using the GMM in estimating the risk premium, and that the model specification test based on the SDF is also as powerful as the one based on the beta method using the GMM. They argue that Kan and Zhou's (1999) conclusion is inappropriate in that they ignored the risk premium measures in the two methods not identical and that they assumed the factor is predetermined. Jagannathan et al. (2002) consider explicitly the additional moments about the SDF method and the transformation between the risk premium parameters in order to make a fair comparison between the two methods.

Let F_t and R_t be a vector of K risk factors and a vector of N test asset returns respectively. The pricing model can be denoted as

$$E[R_t] = \gamma \mathbf{1} + B\Gamma, \quad (10)$$

where γ is the nonzero return of the zero-beta asset, $\mathbf{1}$ is an $(N \times 1)$ vector of ones, R_t is an $(N \times 1)$ vector of raw returns for the test assets, B is the factor loadings of the following time-series regression, $R_t = \varphi + Bf_t + \varepsilon_t$, and the error term ε_t has zero

mean and uncorrelated with the factors f_t . Then $\varphi = \gamma 1 + B(\Gamma - \mu_f)$, where μ_f is the mean of the factors. Using this restriction, the following restrictions can be obtained:

$$R_t = (\gamma 1 + B\Gamma - B\mu_f) + Bf_t + \varepsilon_t \quad (11)$$

$$E[\varepsilon_t] = 0 \quad (12)$$

$$E[\varepsilon_t f_t'] = 0. \quad (13)$$

Now, the following $N + NK + K$ moment conditions can be made:

$$E[R_t - \gamma 1 - B(\Gamma - \mu_f + f_t)] = 0 \quad (14)$$

$$E[(R_t - \gamma 1 - B(\Gamma - \mu_f + f_t))f_t'] = 0 \quad (15)$$

$$E[f_t - \mu_f] = 0. \quad (16)$$

Then, the unknown parameters to be estimated are γ , Γ , B , and μ_f . When the factor is the return on a portfolio of traded assets, it implies $\Gamma = \mu_f$ and the moment conditions are reduced as follows:

$$E[R_t - \gamma 1 - Bf_t] = 0 \quad (17)$$

$$E[(R_t - \gamma 1 - Bf_t)f_t'] = 0. \quad (18)$$

The moment conditions for the SDF representation can be derived from the linear asset pricing model. Using $E[R_t] = \gamma 1 + B\Gamma$ and $B = \Sigma_{Rf} \Sigma_{ff}^{-1}$, the following equation can be obtained:

$$E \left[R_t \left(\frac{1}{\gamma} + \frac{\Gamma' \Sigma_{ff}^{-1} \mu_f}{\gamma} - \frac{\Gamma' \Sigma_{ff}^{-1}}{\gamma} f_t \right) \right] = 1. \quad (19)$$

Then, the following SDF representation can be produced:

$$E[R_t m_t] = 1, \quad (20)$$

where $m_t = \lambda' x_t$, $\lambda = [\lambda_0 \lambda_1 \dots \lambda_k]'$, and $x_t = [1 f_t']'$. Using these moment conditions, the parameters can be estimated, and the specification test can be implemented. The closed-form solution is known as follows:

$$\hat{\lambda} = (\hat{D}\hat{V}^{-1}\hat{D})\hat{D}\hat{V}^{-1}\bar{r}, \quad (21)$$

where $\hat{D} = (1/T) \sum_{t=1}^T R_t x_t'$, $\bar{r} = (1/T) \sum_{t=1}^T r_t$, and $\hat{V}$ is a consistent estimator in Eq. 5.

The moment conditions for the conditional asset pricing are provided by Harvey (1989, 1991), Ferson and Harvey (1993), Jagannathan and Wang (1996), and Cochrane (1996).

63.3.4 The Hansen-Jagannathan Distance

Hansen and Jagannathan (1997) develop a distance metric which is referred as the Hansen-Jagannathan (HJ)-distance. They propose how to measure the distance between a true stochastic discount factor and the implied stochastic discount factor. The HJ-distance can be interpreted as the normalized maximum pricing error of the model for the test assets. If the implied stochastic discount factor from the model is included in the set of true stochastic discount factor, then the HJ-distance is zero and the pricing error is none. To compute the HJ-distance, we can use simply the inverse of the second moment of the test asset returns as the weighting matrix, $E(R_t R_t')^{-1}$. They also provide the distribution of the HJ-distance.

The HJ-distance is defined as follows:

$$\delta = \text{Min} \|y_t - m_t\| \text{ with respect to } m_t, \quad (22)$$

where $E(m_t R_t) = 1$, R_t is the raw returns, y_t is a candidate SDF, and $m_t \in M_t$, where M_t is the set of true SDFs. The measure of distance is the usual norm. The minimization problem can be written as the Lagrangian minimization problem

$$\delta^2 = \text{Min}_{m \in L^2, m \geq 0} \sup_{\tau \in R^n} \left\{ E(y_t - m_t)^2 - 2\tau' [E(m_t R_t) - 1] \right\}, \quad (23)$$

where δ is the minimum distance from the candidate SDF to the true pricing kernel and L^2 is the space of all random variables with finite second moments. Hansen and Jagannathan (1997) show the following relation:

$$y_t - m_t^* = \tau^{*'} R_t, \quad (24)$$

where m_t^* and τ^* are the solutions for the Lagrangian minimization problem, and $\tau^* = E(R_t R_t')^{-1} E(y_t R_t - 1)$. Therefore, the HJ-distance is

$$\delta = \|y_t - m_t^*\| = \left[\tau^{*'} E(R_t R_t') \tau^* \right]^{\frac{1}{2}}. \quad (25)$$

Using $\tau^* = E(R_t R_t')^{-1} E(y_t R_t - 1)$, the following equation can be produced:

$$\delta = \left[E(y_t R_t - 1)' E(R_t R_t')^{-1} E(y_t R_t - 1) \right]^{\frac{1}{2}}, \quad (26)$$

where the pricing errors $g_t(\theta)$ can be defined as $E(y_t R_t - 1)$ here. Equation 26 is exactly the same as $\sqrt{Q}$ of Eq. 4 with the weighting matrix $E(R_t R_t')^{-1}$. The sample analog can be as follows:

$$\hat{\delta} = [g_T(\theta)]' \hat{G}^{-1} [g_T(\theta)], \quad (27)$$

where $\widehat{G} = (1/T) \sum_{t=1}^T R_t T'_t$, and $G = E(R_t R'_t)$.

The HJ-distance means the smallest distance between the candidate SDF and the true SDF. It can also be interpreted as the maximum pricing error of the considered test assets. Consider the following equations:

$$\begin{aligned} |E(y_t R_{it}) - E(m_t^* R_{it})| &= |E[(y_t - m_t^*) R_{it}]| \\ &\leq E[(y_t - m_t^*)^2]^{\frac{1}{2}} E[R_{it}^2]^{\frac{1}{2}}. \end{aligned} \quad (28)$$

The Cauchy-Schwartz inequality is applied to the last term. Then, Equation 25 can be rewritten as follows:

$$\delta = \|y_t - m_t^*\| \geq \frac{|E(y_t R_{it}) - E(m_t^* R_{it})|}{\|R_{it}\|}. \quad (29)$$

Therefore, the HJ-distance can also be interpreted as the normalized maximum pricing error for the considered test assets.

We estimate a vector of the unknown parameters θ and conduct the model specification test using the typical GMM procedure as follows:

$$\widehat{\theta}^{HJ} = \arg \min_{\theta} [g_T(\theta)]' \widehat{G}^{-1} [g_T(\theta)]. \quad (30)$$

The distribution of $\widehat{\theta}^{HJ}$ is the same as in Eq. 6. However, the distribution of the model specification test statistic, which is T times the square of the HJ-distance, is not equal to that of the Hansen's J -statistic. Jagannathan and Wang (1996) suggest that the distribution of $T \left[g_T(\widehat{\theta}^{HJ}) \right]^2$ is a linear combination of $\chi^2(1)$ variables. The distribution for the null hypothesis that the HJ-distance δ is zero is given by

$$T \left[\delta(\widehat{\theta}^{HJ}) \right]^2 \xrightarrow{d} \sum_{j=1}^{N-K-1} \varphi_j \pi_j \text{ as } T \rightarrow \infty, \quad (31)$$

where π_j is a random variable that follows $\chi^2(1)$, and φ_j is the nonzero eigenvalues of the following matrix:

$$\begin{aligned} \Lambda = & V^{1/2} G^{-1/2} \left(I - \left(G^{-1/2} \right)' D \left(D' G^{-1} D \right)^{-1} D' G^{-1/2} \right) \\ & \left(G^{-1/2} \right)' \left(V^{1/2} \right)', \end{aligned} \quad (32)$$

where $V^{1/2}$ and $G^{1/2}$ are the upper triangular matrices in the Cholesky decompositions of V and G . Jagannathan and Wang (1996) suggest that the empirical p -value can be obtained by comparing $T \left[\delta(\widehat{\theta}^{HJ}) \right]^2$ to $\sum_{j=1}^{N-K-1} \varphi_j \pi_j$ which is computed as

random draws from a $\chi^2(1)$ distribution. The empirical p -value of the HJ-distance can be computed from the following function:

$$\text{Empirical } p\text{-value} = \frac{1}{I} \sum_{i=1}^I f(x_i), \quad (33)$$

where $x_i = \left(\sum_{j=1}^{n-k-1} \varphi_j \pi_{ji} \right) - T \left[g_T(\hat{\theta}^{HJ}) \right]^2$, $f(x_i)$ is the function which returns 1 if $x_i \geq 0$ and 0 otherwise. Jagannathan and Wang (1996) set I to 5000 in their paper. According to Ahn and Gadarowski (2004), tests based on this empirical p -value reject too often the correct models, since the computation of the empirical p -value depends on the eigenvalues from the estimate of Λ and $\hat{V}$ has the poor finite sample property.

63.3.5 An Equality Test for the HJ-Distance

Kan and Robotti (2009) develop the method to test the HJ-distance equality of the two competing linear asset pricing models. They analyze the case of both the nested and nonnested models whether the HJ-distances of the two competing models are equal. This section introduces the general case when both nonnested models are misspecified and the both stochastic discount factors are not equal.

Let y_1 and y_2 be two competing models. δ_1 and δ_2 of Eq. 22 in the two models are nonzero. All subscript means the model type. Let $d_t = q_{1t} - q_{2t}$, where $q_{1t} = 2u_{1t}y_{1t} - u_{1t}^2 + \delta_1^2$, $q_{2t} = 2u_{2t}y_{2t} - u_{2t}^2 + \delta_2^2$ with $u_{1t} = g_1' G^{-1} R_t$ and $u_{2t} = g_2' G^{-1} R_t$. Then

$$\sqrt{T} \left[\hat{\delta}_1^2 - \hat{\delta}_2^2 - (\delta_1^2 - \delta_2^2) \right] \xrightarrow{d} N(0, v_d), \quad (34)$$

where $v_d = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]$ and $d_t = 2u_{1t}y_{1t} - u_{1t}^2 - 2u_{2t}y_{2t} + u_{2t}^2$. Under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2 \neq 0$,

$$\sqrt{T} \left(\hat{\delta}_1^2 - \hat{\delta}_2^2 \right) \xrightarrow{d} N(0, v_d) \quad (35)$$

63.4 Cross-Sectional Tests

63.4.1 A Risk Premia Estimation Through Two-Pass Regressions

Traditional asset pricing theories such as the CAPM (Sharpe 1964; Lintner 1965; Black 1972), the APT (Ross 1976), the Intertemporal CAPM (Merton 1973), and the Consumption CAPM (Breedon 1979) imply that the difference in expected returns

across assets should be explained by their covariances with (systematic) risk factors or their factor loadings. When asset returns are generated from a K -factor model, their expected returns are represented by a linear combination of the factor loadings or betas, as in Eq. 2,

$$E(r_{it}) = \gamma' \beta_i \text{ for } i = 1, 2, \dots, N. \quad (36)$$

Since the factor loadings, β_i , are not directly observable, it is natural to test the cross-sectional relation between expected returns and factor loadings in two stages. In the first stage, beta estimates are obtained from time-series regressions of Eq. 1,

$$r_{it} = R_{it} - R_{ft} = \alpha_i + \beta_i' F_t + \varepsilon_{it}, E(\varepsilon_{it}|F_t) = 0. \quad (37)$$

In practice, beta estimates in the first stage regression are computed using the OLS and are given as follows:

$$\widehat{\beta}_i = \widehat{Var}(F)^{-1} \widehat{Cov}(F, r_i), \quad (38)$$

where $\widehat{Var}(F) = (1/T) \sum (F_t - \bar{F})'(F_t - \bar{F})$ and $\widehat{Cov}(F, r_i) = (1/T) \sum (F_t - \bar{F})(r_{it} - \bar{r}_i)$. In the second stage, one runs a cross-sectional regression (CSR) of excess average returns of N assets on their estimated betas obtained from the first stage,

$$\bar{r} = \widehat{\beta} \gamma + e, \quad (39)$$

where $\bar{r}$ is a $(N \times 1)$ average returns, $\widehat{\beta} = [\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_N]'$ is an $(N \times K)$ beta estimates matrix, and e is an $(N \times 1)$ residual returns with mean zero and covariance matrix $\mathbf{Var}(e) = \Sigma_e$.

Under the classical linear model assumptions, the OLS risk premia estimates of Eq. 39 is given as

$$\widehat{\gamma}_{OLS} = (\widehat{\beta}' \widehat{\beta})^{-1} \widehat{\beta}' \bar{r}. \quad (40)$$

However, obtaining the correct distribution of $\widehat{\gamma}_{OLS}$ in finite samples is not so trivial, since the explanatory variables are measured with error and the least squares estimates of Eq. 40 are subject to the errors-in-variables (EIV) bias. If we ignore the EIV bias and regard $\widehat{\beta}$ as the true beta variable β for a while, then the distribution of the OLS estimator of Eq. 40 is

$$\widehat{\gamma}_{OLS} \sim N\left(\gamma, (\beta' \beta)^{-1} \beta' \mathbf{Var}(e) \beta (\beta' \beta)^{-1}\right) \quad (41)$$

Since $\mathbf{Var}(e) = \mathbf{Var}(\bar{r} - \beta \gamma) = \mathbf{Var}(\alpha + \beta \bar{F} + \bar{\varepsilon} - \beta \gamma) = \mathbf{Var}[\beta(\bar{F} - \gamma) + \bar{\varepsilon}]$ the error covariance of the cross-sectional regression in (39) is

$$\text{Var}(e) = \frac{1}{T}(\beta \Sigma_F \beta' + \Sigma_e), \quad (42)$$

where Σ_F is the factor covariance and Σ_e is the error covariance in the time-series regression. Therefore, Equation 41 can be rewritten as

$$\hat{\gamma}_{OLS} \sim N\left(\gamma, \frac{1}{T} \left[(\beta' \beta)^{-1} \beta' \Sigma_e \beta (\beta' \beta)^{-1} + \Sigma_F \right] \right). \quad (43)$$

See also Cochran (2004). For statistical inference, each variable in Eq. 43 can be replaced with its time-series sample counterpart.

63.4.2 The GLS Estimation of Risk Premia

The OLS risk premia estimation in the previous section assumes that the residuals are cross-sectionally uncorrelated. However, since this assumption could be too strong, risk premia can be estimated by the generalized least squares (GLS) in which residuals are weighted in a way to ease heteroskedasticity. The GLS estimate of risk premia is¹

$$\begin{aligned} \hat{\gamma}_{GLS} &= \left[\beta' \text{Var}(e)^{-1} \beta \right]^{-1} \beta' \text{Var}(e)^{-1} \bar{r} \\ &= \left[\beta' (\beta \Sigma_F \beta' + \Sigma_e)^{-1} \beta \right]^{-1} \beta' (\beta \Sigma_F \beta' + \Sigma_e)^{-1} \bar{r} \\ &= (\beta' \Sigma_e^{-1} \beta)^{-1} \beta' \Sigma_e^{-1} \bar{r}. \end{aligned} \quad (44)$$

The distribution of the GLS estimator of Eq. 44 is

$$\hat{\gamma}_{GLS} \sim N\left(\gamma, \frac{1}{T(\beta' \Sigma_e^{-1} \beta + \Sigma_F)}\right). \quad (45)$$

63.4.3 The Errors-in-Variables Problem

63.4.3.1 Corrections for the Standard Errors

Previous discussions simply ignore the fact that beta estimates obtained in the first pass are not true betas but are measured with error (Shanken 1992). It is well known that the beta estimate in time-series regression is T -consistent (as T goes to infinity, the estimation error of beta estimates goes to zero). In other words, as the number of

¹Let $\mathbf{A} = \mathbf{I} + \beta' \Sigma_e^{-1} \beta \Sigma_F$. Then, $\hat{\gamma}_{GLS} = \left[\mathbf{A} \beta' (\beta \Sigma_F \beta' + \Sigma_e)^{-1} \beta \right]^{-1} \mathbf{A} \beta' (\beta \Sigma_F \beta' + \Sigma_e)^{-1} \bar{r}$. Since, $\mathbf{A} \beta' = \beta' \Sigma_e^{-1} (\Sigma_e + \beta \Sigma_F \beta')$, $\hat{\gamma}_{GLS} = (\beta' \Sigma_e^{-1} \beta)^{-1} \beta' \Sigma_e^{-1} \bar{r}$.

time-series observations (T) goes to infinity and they are stationary, the beta estimate converges to its true beta. As long as we obtain T -consistent beta estimates, the CSR least squares estimates of risk premia are N -consistent if the number of test assets in the CSR (N) is large. However, since return observations show nonstationarity (see Kim and Kon (1996, 1999)) and thus it is difficult to increase T in estimating betas, it would be hard to obtain T -consistent beta estimates, and accordingly, the least squares estimates of Eqs. 40 and 44 and their standard errors are not N -consistent and are biased.

Shanken (1992) suggests a method of correcting for the bias of the standard errors of the CSR least squares estimates in the two-pass methodology. Assuming the model given in Equation 37 is correct, Shanken (1992) provides the correction for the variance of the CSR risk premia estimates as follows:

$$Var(\hat{\gamma}_{OLS}) = \frac{1}{T} \left[(1 + \gamma' \Sigma_F^{-1} \gamma) (\beta' \beta)^{-1} \beta' \Sigma_e \beta (\beta' \beta)^{-1} + \Sigma_F \right] \quad (46)$$

$$Var(\hat{\gamma}_{GLS}) = \frac{1}{T} \left[(1 + \gamma' \Sigma_F^{-1} \gamma) (\beta' \Sigma_e^{-1} \beta)^{-1} + \Sigma_F \right]. \quad (47)$$

Notice that t -values of the CSR estimates tend to be overstated due to the measurement errors of the explanatory variables $\hat{\beta}$. Since the corrected variance of (46) or (47) is increased by adding the correction term $\gamma' \Sigma_F^{-1} \gamma$, the above correction attempts to lessen the EIV bias in the standard error of the estimates.

63.4.3.2 Corrections for the Risk Premium Estimate

The Shanken correction in (46) or (47) can be applied only to make a correction for the EIV bias of the standard errors of the CSR risk premia estimates, not of the estimates themselves (Kim 1995). Instead of correcting the variance as in Shanken (1992), Kim (1995) provides a direct correction for the EIV bias of the CSR market risk premium estimate within the Fama and MacBeth methodology. Suppose we have the following CSR model to be estimated at time t with an estimated beta variable and a firm characteristic variable (e.g., firm size or book to market) measured without error at $t - 1$,

$$R_t = \gamma_{0t} + \gamma_{1t} \hat{\beta}_{t-1} + \gamma_{2t} V_{t-1} + \varepsilon_t. \quad (48)$$

In the above equation, the beta variable is measured with error. The beta variable is typically obtained from the time-series regression (the first pass) of a firm's returns on the market returns by using return observations from $t - T$ to $t - 1$. When the beta variable is measured error as in Eq. 48, it is well known that the EIV problem is present. Thus, the EIV problem occurs in the two-pass methodology. The least squares estimation of the CSR in Eq. 48 results in the underestimation of the role of market beta and the overestimation of the role of the firm characteristic variables.

To make correction for the EIV problem, Kim (1995) provides an EIV corrected estimators for the CSR coefficients of Eq. 48, $\hat{\gamma}_{0t}$, $\hat{\gamma}_{1t}$, and $\hat{\gamma}_{2t}$, which are N -consistent.

$$\hat{\gamma}_{1t} = C_{1t} \hat{\gamma}_{1t}^{LS} \quad (49)$$

$$\hat{\gamma}_{2t} = C_{2t} \hat{\gamma}_{2t}^{LS} \quad (50)$$

$$\hat{\gamma}_{0t} = \hat{\gamma}_{0t}^{LS} - (C_{1t} - 1) \hat{\gamma}_{1t}^{LS} \bar{\beta}_{t-1} - (C_{2t} - 1) \hat{\gamma}_{2t}^{LS} \bar{V}_{t-1} \quad (51)$$

where

$$C_{1t} = \frac{1}{\left[1 - \frac{1}{2} \left(\frac{m_{RR} (1 - \hat{\rho}_{RV}^2)}{\delta_t m_{\hat{\beta}\hat{\beta}} (1 - \hat{\rho}_{\beta V}^2)} + 1 \right) (1 - (1 - Q)^{\frac{1}{2}}) \right]} \geq 1$$

$$Q = \frac{1 - \hat{\rho}_{R\beta V}^2}{\left[\frac{1}{2} + \frac{1}{4} \left(\frac{\delta_t m_{\hat{\beta}\hat{\beta}} (1 - \hat{\rho}_{\beta V}^2)}{m_{RR} (1 - \hat{\rho}_{RV}^2)} + \frac{m_{RR} (1 - \hat{\rho}_{RV}^2)}{\delta_t m_{\hat{\beta}\hat{\beta}} (1 - \hat{\rho}_{\beta V}^2)} \right) \right]}$$

$$C_{2t} = 1 - \frac{(C_{1t} - 1) \left\{ (\hat{\rho}_{R\beta} \hat{\rho}_{\beta V} / \hat{\rho}_{RV}) - \hat{\rho}_{\beta}^2 \right\}}{1 - (\hat{\rho}_{R\beta} \hat{\rho}_{\beta V} / \hat{\rho}_{RV})}$$

and $\delta_t = T \hat{\sigma}_{m,t-1}^2$ where $\hat{\sigma}_{m,t-1}^2$ is the MLE of the market variance computed using market returns from $t - T$ to $t - 1$, $m_{xy} = (1/N)(x - \bar{x})' \hat{\Sigma}_s^{-1} (y - \bar{y})$, $\bar{x} = \Sigma_{i=1}^N \Sigma_{j=1}^N w_{ij} x_i / \Sigma_{i=1}^N \Sigma_{j=1}^N w_{ij}$, w_{ij} is the (i,j) -th element of $\hat{\Sigma}_\varepsilon^{-1}$, $\hat{\rho}_{xy} = m_{xy} / (m_{xx} m_{yy})^{1/2}$ and $\hat{\rho}_{R\beta V}^2 = (\hat{\rho}_{R\beta} \hat{\rho}_{\beta V} / \hat{\rho}_{RV})^2 / \left\{ (1 - \hat{\rho}_{\beta V}^2) (1 - \hat{\rho}_{RV}^2) \right\}$. Here, $\hat{\gamma}_t^{LS}$ is the least squares estimate. Note that $m_{\cdot \cdot}$ is the cross-sectional sample second moment and $\rho_{\cdot \cdot}$ is thus the cross-sectional correlation coefficient.

Since the EIV-bias correction term C_{1t} is greater than one, this term makes correction for the underestimation of the least squares estimate, $\hat{\gamma}_{1t}^{LS}$. Also, since the term C_{2t} is usually less than one, this term makes correction for the overestimation of the least squares estimate of the coefficient on the firm characteristic variable such as firm size and book to market, $\hat{\gamma}_{2t}^{LS}$. Following the seminal works of Black et al. (1972) and Fama and MacBeth (1973), portfolio grouping procedures are prevalent to lessen the EIV problem. Aside the fact that the portfolio grouping procedure does not completely remove the problem, however, it significantly reduces the sample size (N), henceforth weakens the power of the tests. In this regard, a direction correction for the EIV problem like Kim (1995) is important for the use of all individual assets as test assets.

63.5 Summary and Concluding Remarks

This chapter reviews the methodologies of testing asset pricing models. We provide some explanations for two testing methodologies dominantly used in the literature: time-series regression tests and cross-sectional regression tests. We explain individual t -test and the joint F -test by Gibbons et al. (1989) which are used in time-series tests. We also explain the two-pass test methodology that is used in cross-sectional test, and discuss the errors-in-variables problem which occurs inevitably in the two-pass methodology. Moreover, we provide some explanations of tests based on the GMM estimation and tests based on the Hansen and Jagannathan (1997) distance.

References

- Ahn, S.C., and C. Gadarowski. 2004. Small sample properties of the GMM specification test based on the Hansen-Jagannathan distance. *Journal of Empirical Finance* 11: 109–132.
- Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.
- Black, F. 1972. Capital market equilibrium with restricted borrowing. *Journal of Business* 45: 444–455.
- Black, F., M.C. Jensen, and M.S. Scholes. 1972. The capital asset pricing model: Some empirical tests. In *Studies in the theory of capital markets*. Praeger Publishers Inc.
- Breeden, D. 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics* 7: 265–296.
- Cochrane, J.H. 1996. A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy* 104: 572–621.
- . 2004. *Asset Pricing*. Princeton University Press.
- Fama, E.F., and K.R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56.
- Fama, E.F., and J.D. MacBeth. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 81: 607–636.
- Ferson, W.E., and S.R. Foerster. 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. *Journal of Financial Economics* 36: 29–55.
- Ferson, W.E., and C.R. Harvey. 1993. The risk and predictability of international equity returns. *Review of Financial Studies* 6: 527–566.
- Gibbons, M.R., S.A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* 57: 1121–1152.
- Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.
- Hansen, L.P., and R. Jagannathan. 1997. Assessing specification errors in stochastic discount factor models. *Journal of Finance* 62: 557–590.
- Harvey, C.R. 1989. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics* 24: 289–317.
- . 1991. The world price of covariance risk. *Journal of Finance* 46: 111–157.
- Hayashi, F. 2000. *Econometrics*. Princeton University Press.
- Jagannathan, R., G. Skoulakis, and Z. Wang. 2002. Generalized method of moments: Applications in finance. *Journal of Business & Economic Statistics* 20: 470–481.
- Jagannathan, R., and Z. Wang. 1996. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51: 3–53.

- Kan, R., and C. Robotti. 2009. Model comparison using the Hansen-Jagannathan distance. *Review of Financial Studies* 22: 3449–3490.
- Kan, R., and G. Zhou. 1999. A critique of the stochastic discount factor methodology. *Journal of Finance* 54: 1221–1248.
- Kim, D. 1995. The errors-in-variables problem in the cross-section of expected stock returns. *Journal of Finance* 50: 1605–1634.
- . 2011. Asset pricing models in the Korean Stock Markets: A review for the Period of 1980 ~ 2009. *Asian Review of Financial Research* 24: 167–229.
- Kim, D., and S. Kon. 1996. Sequential parameter nonstationarity in stock market returns. *Review of Quantitative Finance and Accounting* 6: 103–131.
- . 1999. Structural change and time dependence in models of stock returns. *Journal of Empirical Finance* 6: 283–308.
- Ledoit, O., and M. Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10: 603–621.
- Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867–887.
- Lintner, J. 1965. The valuation of risky assets and the selection of risky investment in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–37.
- Newey, W.K., and K.D. West. 1987. Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28: 777–787.
- Ren, Y., and K. Shimotsu. 2009. Improvement in finite sample properties of the Hansen-Jagannathan distance test. *Journal of Empirical Finance* 16: 483–506.
- Ross, S.A. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13: 341–360.
- Shanken, J. 1992. On estimation of beta pricing models. *Review of Financial Studies* 5: 1–33.
- Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- Windmeijer, F. 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126: 25–51.



Unified Model Arbitrage-Free Term Structure of Flow Risks

64

Thomas S. Y. Ho and Sang Bin Lee

Abstract

Interest movement models are important to financial modeling because they can be used for valuing any financial instruments whose values are affected by interest rate movements. Specifically, we can classify the interest rate movement models into two categories: equilibrium models and no-arbitrage models. The equilibrium models emphasize the equilibrium concept. However, the no-arbitrage models argue that the term structure movements should satisfy the no-arbitrage condition. The arbitrage-free interest rate model is an extension of the Black-Scholes model to value interest rate derivatives. The model valuation is assured to be consistent with the observed yield curve in valuing interest rate derivatives and providing accurate pricing of interest rate contingent claims. Therefore, it is widely used for portfolio management and other capital market activities.

Keywords

Stock risk driver · Flow risk driver · Unified model · Interest model · Credit valuation model · Liquidity valuation model · Energy valuation model · Inflation contingent claims valuation model · State-time-dependent volatilities · Regime switching · Multi-movement model · Arbitrage-free condition · Term structure of interest rates · Term structure movements

The authors would like to thank the participants at Financial Innovation Conference, Owen School, Vanderbilt University 2008 for their comments. We would also like to thank Eric A Anderson and Charles Fenner for their insights into energy contracts. We are responsible for any errors in the chapter.

T. S. Y. Ho (✉)

Thomas Ho Company, Ltd., New York, NY, USA

e-mail: Tom.Ho@ThomasHo.com

S. B. Lee

Hanyang University, Seoul, South Korea

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_24

1485

64.1 Introduction

Economic quantities are classified by stocks and flows. The former are values at a point in time and the latter are quantities measured per unit of time. This classification of economic quantities is widely used in accounting and econometric models, where balance sheet items and capital are considered stocks and income statement items and transactions are flows. The concept of stocks and flows is also applicable to the understanding of risks in the financial markets, where risk drivers can also be dichotomized into stocks and flows.

Stock risk drivers are exemplified by equities, commodities, and firm values. They are the underlying risk drivers to many financial instruments, such as equity options and commodity futures. At any time, stock risk drivers have observable values that follow some stochastic processes which affect the values of their contingent claims. Flow risk drivers such as the interest rates and inflation rates also have their stochastic movements. However, their value is measured per unit of time. And therefore, their contingent claims are generally affected by the accumulated flow risks over time. Therefore, a flow risk driver must be related to a term structure of risks, such as the term structure of interest rates and the stochastic movements of the term structure, and not just the stochastic movements of the rate, are important to the understanding of flow risks.

The Black-Scholes model (1973) is a relative valuation model that values contingent claims on stock risk drivers. It assumes that in a perfect capital market the stochastic movements of a derivative and its underlying security are instantaneously perfectly correlated. This characteristic of a derivative enables the market participants to form dynamic hedging strategies to construct a risk-free portfolio of derivatives and their underlying securities and that in turn leads to the risk-neutral valuation of the derivative relative to the observed underlying security price. Furthermore, the model suggests that benchmark derivative prices can be used to determine the implied volatility of the underlying security returns, and the implied volatility can be used to price other derivatives based on the same underlying securities. That is, the relative valuation model can be defined by the underlying security and a set of benchmark contingent claims or options on the security. This fundamental insight has been elaborated in the continuous-time model context. See Harrison and Kreps (1979).

The Ho-Lee model (1986) is also a relative valuation model based on the Black-Scholes paradigm. It is a valuation model of contingent claims on flow risk drivers, as opposed to the stock risk drivers. The model specifies arbitrage-free movements of flow risk drivers to value a broad range of financial instruments. The term “arbitrage-free” is defined as the absence of arbitrage opportunity among the contingent claims and the observed term structure of securities. This definition is consistent with that of the Ho-Lee model paradigm. Dynamic hedging between the term structure of risks and its derivatives is used to derive the risk-neutral valuation of the derivatives, and the benchmark derivatives with embedded options are used to determine the implied volatilities of the term structure movements. Details

of these concepts described in a continuous-time model context can be found in Rebonato (1998).

The model has generated extensive research in alternative specifications of term structure movements of interest rates and applications in capital markets. The purpose of this entry is to present the Ho-Lee model as a unified model, in the sense that (1) it is a contingent claim valuation model on flow risk drivers, not just a model confined to interest rates, and (2) it is a model that combines multiple risk drivers to determine the contingent claims value, not just an analysis of contingent claims on a particular risk driver, isolated from other risk drivers.

The unified model is an important contribution to financial research because the valuation framework applies to financial instruments that are prevalent in the capital markets; they can be found in trading positions, investment portfolios, balance sheets of financial institutions, energy sectors, inflation-linked financial products in pension plans, and other financial sectors. The unified model is also important in providing a framework to value and hence to analyze financial instruments with multiple sources of risks, a topic particularly important in enterprise risk management where risk managers have the daunting task of rolling upmarket risks, credit risks, and liquidity risks into a coherent framework.

The financial crisis of 2008 clearly shows the importance of the application of a unified model. First, the risks of securities such as asset-backed securities and collateralized bond obligations and the credit derivative swaps of super senior CDO tranches are option embedded contingent claims on credit and market risks. Much research has been devoted to analyzing interest rate options embedded in financial instruments. However, much less research has been devoted to value credit options and liquidity options, and scant attention has been devoted to the combined effects of these three major risk factors on financial instruments. For example, the observed increased default rate, lack of liquidity in the mortgage loan and mortgage-backed securities markets, and slowing of refinancing of mortgages are not isolated events but are consequences of the combined effect of credit, interest rate, and liquidity risks on the mortgage value and the financial markets. Standard approaches in analyzing these securities would significantly measure the risks erroneously resulting in significant losses in the complex securities.

Second, the risks of enterprises should not be analyzed in isolation. Basel II has advocated that risks should be analyzed by separate departments. When risks are managed in silos, enterprise risk management, regulatory monitoring, and supervision processes would fail to provide a coherent framework to manage the risks in the financial system. Financial institutions fail to manage the combined effect of the flow risks in the financial crisis. Third, the economic capital of a financial institution should be recognized as a contingent claim of the market, credit, and liquidity risk and it has a significant embedded option on these risk drivers. It cannot be measured simply as the asset value net of the liability value as it is commonly reported. The option embedded in the economic capital may lead to unanticipated consequences in the “deleveraging” of a balance sheet. This chapter will relate the unified model to these issues, which are significant causes of the financial crisis.

The chapter proceeds as follows. In Sect. 64.3.2 we identify the differences between a stock risk driver and a flow risk driver. And we present the Ho-Lee model (Ho and Lee 1986) as a unified model. Section 64.3.3 provides five examples of contingent claims on flow risk drivers. Section 64.3.4 describes the unified model in valuing contingent claims on multiple flow risk drivers, where the term structure of the risk drivers has multiple movements. The section also describes the modeling of these movements. Section 64.3.5 describes the implications of the unified model in enterprise risk management. And finally, the conclusion follows.

64.2 The Black-Scholes Model and the Ho-Lee Model

In this section, we use a binomial lattice model to illustrate the Black-Scholes model in a way comparable to the Ho-Lee model. Let $s(n, i)$ and $C(n, i)$ be the underlying security and the call option, respectively, at time n and state i . Let the option time to expiration and strike be T and X , respectively. Let $S(n)$ be the forward price of the security, where

$$S(n) = Se^{nr} \quad (1)$$

The stochastic price $s(n, i)$ is defined by Eq. (2) with the binomial probability of 0.5 and with a given risk-free rate r :

$$s(n, i) = S(n)(2/(1 + \delta))^n \delta^i \quad (2)$$

Then, the Black-Scholes model can be expressed as a set of recursive equations.

$$C(n - 1, i) = 0.5e^{-r}(C(n, i + 1) + C(n, i)) \text{ for } i = 0, \dots, n - 1 \quad (3)$$

with the terminal condition for a call option, as an illustration

$$C(T, i) = \max(S(T, i) - X, 0) \text{ for } i = 0, 1, 2, \dots, T \quad (4)$$

The value of the call option is $C = C(0, 0)$. The binomial movement of the security S ensures that the option price is the observed stock price S , when the strike price is zero, satisfying the arbitrage-free condition. δ determines the local volatility of the underlying security. When C is observable in the market, then the option price can be used to determine δ , and thus specifying the implied volatility of the underlying security, which in turn can be used to value other derivatives of S . The underlying security S is a stock risk driver.

δ determines the volatility of the stock σ . Indeed, we have $\delta = \text{Exp}\left[-2\frac{\sigma}{\sqrt{n}}\right]$, where n is the number of steps in one year. This can be proved as follows. The stock price at

time n and state i is $Se^{nr}\left(\frac{2}{1+\delta}\right)^n \delta^i$. The logarithm of the stock return is $\text{Log}\left[\frac{Se^{nr}\left(\frac{2}{1+\delta}\right)^n \delta^i}{S}\right] = nr + n \text{Log}\frac{2}{1+\delta} + i \text{Log}\delta$ and the variance is $(\log \delta)^2 \frac{1}{2} (1 - \frac{1}{2})n = \sigma^2$.

Proposition 1 The binomial model defined by (1)–(4) converges to the continuous-time Black-Scholes model

Proof First, we show that the binomial distribution approaches the lognormal distribution as $n \rightarrow \infty$.¹

We know that

$$\log\left(\frac{S_T}{S_0}\right) = \mu\tau + \sigma\sqrt{\tau}Z, \text{ where } Z \text{ is a standard normal distribution.} \quad (5)$$

From (2),

$$\log\left(\frac{S_T}{S_0}\right)_n = r\tau + n \log\left(\frac{2}{1+\delta}\right) + j \log(\delta) \quad (6)$$

From the DeMoivre-Laplace theorem, $\lim_{n \rightarrow \infty} \text{Prob}\left[a \leq \frac{j-nq}{\sqrt{nq(1-q)}} \leq b\right] = N(b) - N(a)$.

Since q is 0.5, $j \approx (\sqrt{n}/2)Z + n/2$ where the notation $\approx$ means converge in distribution.

Substituting into (6) gives $\log\left(\frac{S_T}{S_0}\right)_n = r\tau + n \log\left(\frac{2}{1+\delta}\right) + ((\sqrt{n}/2)Z + n/2)\log(\delta)$

Let $\mu\tau = r\tau + n \log\left(\frac{2}{1+\delta}\right) + n/2 \cdot \log(\delta)$ and $\sigma\sqrt{\tau} = \sqrt{n}/2 \cdot \log(\delta)$. then

$$\begin{aligned} N(b) - N(a) &= \lim_{n \rightarrow \infty} \text{Prob}\left[a \leq \frac{j-nq}{\sqrt{nq(1-q)}} \leq b\right] = \lim_{n \rightarrow \infty} \text{Prob}\left[a \leq \frac{j-\frac{n}{2}}{\sqrt{n/4}} \leq b\right] \\ &= \lim_{n \rightarrow \infty} \text{Prob}\left[a \leq \frac{r\tau + n \log\left(\frac{2}{1+\delta}\right) - r\tau - n \log\left(\frac{2}{1+\delta}\right) + \log(\delta)j - \log(\delta) \cdot n/2}{\log(\delta) \cdot \sqrt{n/4}} \leq b\right] \\ &= \lim_{n \rightarrow \infty} \text{Prob}\left[a \leq \frac{\log\left(\frac{S_T}{S_0}\right)_n - \mu\tau}{\sigma\sqrt{\tau}} \leq b\right] \end{aligned}$$

Therefore, $(S_T)_n$ converges in distribution to S_T .

Second, what we have to show is that

¹We assume that τ is 1 for simplicity.

$$c = e^{-r\tau} E[\text{Max}[0, (S_T)_n - K]] - > e^{-r\tau} E[\text{Max}[S_T - K]] \text{ as } n \rightarrow \infty$$

If the binomial probability distribution converges to the lognormal, then its moments must also converge to the moments of the lognormal.

Therefore,

$$c = e^{-r\tau} \sum_{j=0}^n \binom{n}{j} \left(\frac{1}{2}\right)^n \text{Max} \left[0, Se^{r\tau} \left(\frac{2}{1+\delta}\right)^n \delta^j - K \right] \rightarrow SN(d_1) - Ke^{-r\tau} N(d_1 - \sigma\sqrt{\tau})$$

as $n \rightarrow \infty$.

Proposition 2 The stochastic price $S(n, i)$ satisfies the no-arbitrage condition. That is, the risk-neutral present value of the expected value of $S(n, i)$ is the observed price S and the following relationship holds $s(n, i) = 0.5e^{-r}(s(n+1, i+1) + s(n+1, i))$.

Proof The expected value of the stock at period n is $\sum_{i=0}^n Se^{rn} \left(\frac{2}{1+\delta}\right)^n \delta^i \frac{n!}{(n-i)!i!} \frac{1}{2^n} = Se^{rn}$. Therefore, the discounted value of Se^{rn} as of time 0 is S .

Substituting $S(n+1)(2/(1+\delta))^{n+1} \delta^{i+1}$ ($S(n+1)(2/(1+\delta))^{n+1} \delta^i$) into $S(n+1, i+1)$ ($S(n+1, i)$) and simplifying, we can show that the arbitrage-free condition $s(n, i) = 0.5e^{-r}(s(n+1, i+1) + s(n+1, i))$ holds.² QED.

The Ho-Lee model (1986) extends the Black-Scholes derivative valuation model in the perfect capital market context to value contingent claims on a term structure of flow risk drivers. A risk factor $u(n, i)$ is a flow risk driver that follows a stochastic process. There is the concept of “flow,” where a per unit time construct is necessary for describing $u(n, i)$. The symbol “ u ” is used to denote the unified model that applies a broad range of flow risk drivers. A rate index $\rho(n, i)$ is defined as the proportional change of a stochastic quantity over a binomial period, and it is related to the flow risk driver by:

$$u(n, i) = \exp(-\rho(n, i)) \quad (7)$$

A binomial lattice with time n and state i represents the stochastic movements of the risk factor $u(n, i)$ or the rate $\rho(n, i)$. At the initial time ($n = 0$), the market can form the forward factor per unit time period $u(n, i)$ or the forward rate per unit time period $\rho(n)$ where the time to the delivery date is n . As time passes, a path in the binomial lattice is used to represent the realization of events. The forward factor or rate would converge to the realized risk factor and the rate at time n . The forward risk factor and the rate are also related by:

$$U(n) = \exp(-\rho(n)) \quad (8)$$

²It is known that the discrete CRR model converges to the continuous BS model as the number of partitions grows. By the same token, Eqs. (19–22) also converge to the continuous BS model at the same speed.

The term structure of a flow risk driver is defined as a term structure of marketable forward contracts on the flow risk driver. Initially, at time 0, the forward values are $u(n)$ for $n = 1, 2, \dots, T$. Then the Ho-Lee model shows that the stochastic movement of the flow risk driver is given by:

$$u(n, i) = U(n)(2/(1 + \delta^n))\delta^i \quad (9)$$

The contingent claim on the flow risk driver is given by:

$$C(n-1, i) = 0.5 u(n, i)(C(n, i+1) + C(n, i)) \text{ for } i = 0, \dots, n-1 \quad (10)$$

with the terminal condition for a call option, as an illustration,

$$C(T, i) = \max(u(T, i) - X, 0) \text{ for } i = 0, 1, 2, \dots, T \quad (11)$$

Analogous to the Black-Scholes model, when the strike price X is zero and time to expiration is T , in applying the recursive equations, we can retrieve the option

value which equals $e^{-\sum_{i=0}^T U(i)}$, the forward price.³

Lemma Let the forward rate per unit time period be given by $\rho(n)$ for the time to delivery, $n > 0$. The primitive security value is given by.

$$A(n, i) = \frac{\text{Exp}\left[-\sum_{t=0}^{n-1} \rho(t)\right] f(n, i)}{\prod_{t=0}^{n-1} (1 + \delta^t)} \text{ for some real value } 0 < \delta < 1 \quad (12)$$

where $f(n, i)$ is a function of integer n and i and it is uniquely determined by the recursive relation

$$f(m, k) = \delta^{k-1} f(m-1, k-1) + \delta^k f(m-1, k) \quad (13)$$

with boundary conditions.

$$f(m, m) = \delta^{\frac{m(m-1)}{2}}, f(m, 0) = 1, f(m, 1) = \sum_{t=1}^m \delta^{m-t}, \quad (14)$$

³ If we roll back from period T where the value at time T and state i is $u(T, i)$, then we have $e^{-\sum_{i=0}^T U(i)}$ equals the forward price

and

$$f(m, m-1) = \left(\sum_{t=1}^m \delta^{m-t} \right) \delta^{\frac{m(m-1)}{2}}. \quad (15)$$

Proof We prove the lemma by induction. First, we show that Eq. (12) holds for $n = 1$ and $i = 0, 1$. For $n = 1$ and $i = 0$, $A(1, 0) = \frac{\text{Exp}[-\rho(0)]}{2}$ by Eq. (12). But by Eq. (10), we note that $A(1, 0) = \frac{\text{Exp}[-\rho(0)]}{2}$. Similarly, we can show that Eq. (12) holds for $n = 1$ and $i = 1$. Next, we assume that Eq. (12) is correct for all periods up to $n-1$ and for all the states in these periods. The induction proof requires us to show that, given assumption, Eq. (12) also holds for time n and all the states at time n . Consider the pricing of $A(n, i)$. By definition, at time n , the only payoff is at state i and it is \$1. Then at time $n - 1$, two states have positive payoffs. At state i and time $n - 1$, the payoff is.

$\frac{1}{2} \text{Exp}[-\rho(n-1)] \frac{2}{(1+\delta^{n-1})} \delta^i$. Similarly, at state $i-1$ and time $n-1$, the payoff is $\frac{1}{2} \text{Exp}[-\rho(n-1)] \frac{2}{(1+\delta^{n-1})} \delta^{i-1}$. Finally, we use the induction hypothesis. Using primitive securities to price an asset, we have

$$\begin{aligned} & \frac{1}{2} \text{Exp}[-\rho(n-1)] \frac{2}{(1+\delta^{n-1})} \delta^i A(n-1, i) + \frac{1}{2} \text{Exp}[-\rho(n-1)] \\ & \times \frac{2}{(1+\delta^{n-1})} \delta^{i-1} A(n-1, i-1) \end{aligned} \quad (16)$$

Substituting into $A(n-1, i)$ and $A(n-1, i-1)$ by the induction hypothesis, we confirm that Eq. (12) holds.⁴ Q.E.D.

Proposition 3 The model $u(1) - u(4)$ provides an arbitrage-free valuation model for a flow risk driver, given the observed term structure of forward values $U(n)$.

Proof We can show that $\sum_{i=0}^n A(n, i) = \text{Exp} \left[-\sum_{t=0}^{n-1} \rho(t) \right]$, which says that the sum of the primitive securities over the states at time n is equal to the security with \$1 at time n . Specifically, the sum of Arrow-Debreu primitive securities is 1 at time 1 if we assume that the interest rates at every period is zero for simplicity (i.e., $\rho(t) = 0$ for $t \geq 0$). Now, we assume that the sum of Arrow-Debreu primitive securities at time n is 1, which means that the no-arbitrage condition holds at time n . Consider a zero-coupon bond that matures at time $n+1$. The value of the zero-coupon bond at time n is

⁴For a general form of Arrow-Debreu primitive securities, see [Appendix A](#).

$$\sum_{i=0}^n \frac{2\delta^i}{(1+\delta^n)} \frac{f(n,i)}{\prod_{t=0}^{n-1} (1+\delta^t)} = \sum_{i=0}^n \frac{2\delta^i f(n,i)}{\prod_{t=0}^n (1+\delta^t)} = 1 \quad (17)$$

To avoid the arbitrage opportunities, the value of the zero-coupon bond at time 0 is 1. The sum of the Arrow-Debreu primitive securities at time $n+1$ is

$$\sum_{i=0}^{n+1} \frac{f(n+1,i)}{\prod_{t=0}^n (1+\delta^t)} = \sum_{i=0}^n \frac{\delta^i f(n,i)}{\prod_{t=0}^n (1+\delta^t)} + \sum_{i=1}^{n+1} \frac{\delta^{i-1} f(n,i-1)}{\prod_{t=0}^n (1+\delta^t)} = \sum_{i=0}^n \frac{2\delta^i f(n,i)}{\prod_{t=0}^n (1+\delta^t)} \quad (18)$$

Equation (15) is equal to 1 by Eq. (14). Q.E.D.

In comparing Eqs. (23–26) and (7–11), we can see the similarity between the Black-Scholes model and the Ho-Lee model, particularly in their determination of the contingent claim values. In comparing Eqs. 24 and 9, the functional form of the two models is also similar in that it is a product of three factors. The main difference lies on the specification of the forward contracts. The Black-Scholes model specifies the forward price of a stock while the Ho-Lee model specifies the forward contract on a flow entity, where the time unit is defined by the binomial period. Also, compare (25) and (10), the risk-free rate is used in the rollback for the stock risk model and the risk driver is used for the flow risk model. Expressed in the continuous models, a flow risk driver has to be expressed as an “instantaneous rate entity,” such as a money market account.

64.3 Examples of Term Structure of Flow Risk Drivers

64.3.1 Interest Rate Model

An example of the Ho-Lee model is the interest rate model. The discount index $u(n, i)$ is the discount factor for a one binomial period at time n and state i . Specifically, it is the present value of \$1 paid at time $n+1$ at the binomial outcomes in the states i and $i+1$. In the interest rate model, we use the observed yield curve or the discount function $P(n)$ as input to the model. Given the discount function $P(n)$, we can derive the forward values:

$$p(n) = P(n+1)/P(n) \quad (19)$$

Then the stochastic movements of the discount factor are given by:

$$p(n, i) = p(n)(2/(1+\delta^n))\delta^i \quad (20)$$

Heath, Jarrow, and Morton (1992) provide an analogous continuous-time framework to Eq. (20). The δ represents the volatilities and factor $2/(1+\delta^n)$ is called the

convexity term that ensures the yield curve movements are arbitrage free. The valuation of interest rate contingent claims is given by

$$C(n-1, i) = 0.5 p(n, i)(C(n, i+1) + C(n, i)) \text{ for } i = 0, \dots, n-1 \quad (21)$$

with the terminal condition for a call option, as an illustration, given by

$$C(T, i) = \max(p(T, i) - X, 0) \text{ for } i = 0, 1, 2, \dots, T \quad (22)$$

To show that the interest rate model is arbitrage free, we show that the zero-coupon bond price based on the initial forward rates equals that based on the backward substitution methodology. We use the method of induction for the proof. We assume that the initial forward rates $\rho(n)$ where $n = 0, 1, 2, \dots, T$ are given. $\rho(n)$ is the initial forward rate which applies over the period of n to $n+1$. Consider a zero-coupon bond. The face value is assumed to be 1. The price of a zero-coupon bond that matures at period 1 is $e^{-\rho(0)}$. The price of the same bond by backward substitution on the binomial lattice is also $e^{-\rho(0)}$. The price of a zero-coupon bond which matures at period 2 is $e^{-\rho(0)}e^{-\rho(1)}$. The price of the same bond by backward substitution on the binomial lattice is $\frac{1}{2}e^{-\rho(0)}\left(e^{-\rho(1)}\frac{2}{1+\delta}\delta + e^{-\rho(1)}\frac{2}{1+\delta}\right) = e^{-\rho(0)}e^{-\rho(1)}$. Now we assume that the price of a zero-coupon bond that matures at period T is the same regardless of whether we get by backward substitution or from the initial forward rates. The price is $e^{-\rho(0)}e^{-\rho(1)}e^{-\rho(2)}\dots e^{-\rho(T-1)}$. The price of a zero-coupon bond that matures at period $T+1$ is $e^{-\rho(0)}e^{-\rho(1)}e^{-\rho(2)}\dots e^{-\rho(T-1)}e^{-\rho(T)}$ from the initial forward rates. The price of the same bond by backward substitution is also $e^{-\rho(0)}e^{-\rho(1)}e^{-\rho(2)}\dots e^{-\rho(T-1)}e^{-\rho(T)}$, because the price of a zero-coupon bond which matures at $T+1$ is equal to $e^{-\rho(T)}\frac{2\delta}{1+\delta-T}A(T, i)$ where $A(T, i)$ is Arrow-Debreu primitive securities at time T and state i . Adding over i and simplifying, we get $e^{-\rho(0)}e^{-\rho(1)}e^{-\rho(2)}\dots e^{-\rho(T-1)}e^{-\rho(T)}$.

Swaptions can be used to calibrate the interest rate volatility, specified by the parameter δ . The Brace, Gatarek, and Musiela (1997) interest rate model is called a market model as the specification of the model fits both the swap curve and a set of swaption prices exactly. The prevalent use of swaptions in specifying the interest rate model shows the importance of the use of implied volatilities in contingent claims pricing. By way of contrast, Cox, Ingersoll, and Ross (1985) and Vasicek (1977) interest rate models are not relative valuation models in the sense of the Black-Scholes model. These models specify the interest rate risk process to derive the yield curve and they do not take the yield curve as given, using Eq. (19).

64.3.2 Credit Valuation Model

The unified model can be applied to the credit valuation model. Details of a credit model are described in Ho-Lee (2009a, 2009b). The marginal survival factor $s(n, i)$ is the flow risk driver for the credit valuation for a one binomial period at time n and

state i . Specifically, it is the probability for a bond at time n and state i to survive another period, conditional on the bond having survived till time n . In the credit valuation model, we use the observed marginal default rate curve or the survival function $S(n)$ as input to the model. Given the survival function $S(n)$, we can derive the forward values:

$$s(n) = S(n)/S(n-1) \quad (23)$$

Then the stochastic movements of the survival factor are given by:

$$s(n, i) = s(n)(2/(1 + \delta^n))\delta^i \quad (24)$$

The valuation of credit risk contingent claims is given by

$$C(n-1, i) = 0.5 s(n, i)(C(n, i+1) + C(n, i)) \text{ for } i = 0, \dots, n-1 \quad (25)$$

with the terminal condition for a call option, as an illustration

$$C(T, i) = \max(s(T, i) - X, 0) \text{ for } i = 0, 1, 2, \dots, T \quad (26)$$

However, unlike the interest rate model, the survival function and the option of the survival rates of bond is not directly observable in the capital markets. We need to relate the survival function to tradable financial instruments. Credit derivative swaps (CDS) can be used for this purpose. These swaps are actively traded in the market with a term structure of CDS premium observable to the marketplace for many corporate bonds, and even sovereign entities.

The CDS spread of $c(T)$ with a T period tenor is the solution to the set of recursive equations.

$$V(n-1) = p(n)[s(n)V(n) - (1-s(n))(1-R) + c(T)] \text{ for } n = 1, \dots, T \quad (27)$$

with the terminal conditions $V(T) = V(0) = 0$, where R is the recovery rate, $p(n)$ the time value discount factor for one period at time n . Equation (27) says that the seller of the CDS would receive the credit premium of the period $(n-1)$ but would have to pay the loss at default in case the default event occurs during this period. The calculation is conditional to the bond survives till time $(n-1)$. Therefore, given the CDS premiums $c(n)$ for $n = 1, 2, \dots, N$, we can determine the $s(n)$.

In this example, CDSs are used to determine the credit premiums in defining the implied credit premium, a forward-looking estimate of credit risk, as opposed to using the historical defaults experience. However, the method can be used for non-option embedded bonds to estimate the credit premium if the liquidity premium can be appropriately isolated from the credit premium.

The stochastic movements of the CDS curve are different from those of the interest rate curve. While the survival function, like the discount function,

should be monotonically downward sloping and positive, its movement may follow a different mean-reversion process. The specification of the binomial lattice should also reflect such differences. For example, the CDS curve can rise rapidly.

Papers are proposing arbitrage-free valuation methodologies of credit contingent claims such as Duffie (2005), Das et al. (2006), and Longstaff and Rajan (2008). These models assume that the hazard rate follows a mean-reversion process similar to that of the Cox, Ingersoll, and Ross model. And the model is then calibrated to the observed CDS curve. This approach must necessarily calibrate the hazard rate movement model confined to the mean-reversion process. These models cannot separate the calibration of the CDS curve and its volatilities. By way of contrast, this chapter extends the generalized Ho-Lee model to the CDS curve movements that ensures the hazard rate movement is arbitrage free for any given CDS curve. This feature separates the specification of volatilities of the hazard rate from the fitting of the model to the CDS curve. This separation enables the model to have several advantages over other models.

64.3.3 Liquidity Valuation Model

Liquidity is broadly defined as a measure of the performance of the market, the ability to transact any time without moving the market price. A fixed-income security may not have a ready market at any time and a premium is added to the discounting of the cash flows to compensate for such illiquidity. Liquidity premiums vary with time in a stochastic way and they affect the values of fixed-income securities. We have defined the credit spread of a bond by its credit derivative swap curve. But liquidity spread is inseparable from the credit spreads. Lower credit corporate papers tend to be less liquid. For this reason, we can define the liquidity spread to be the discount rate of a bond net of the interest rate and the credit spread. Liquidity spread within the context of our model is defined as the catch-all term, the option-adjusted returns net of time value, and credit risks for a credit paper. In our relative valuation approach, we are not concerned with the modeling of the economics of liquidity. Specifically, we denote the liquidity forward price and the liquidity function to be $l(n)$ and $L(n)$, respectively, then $l(n)$ can be derived from a zero-coupon bond price $V(T)$ with maturity T by the following model:

$$V(T) = P(T)S(T)L(T) \quad \text{for } T = 1, 2, \dots \quad (28)$$

We assume that there is no recovery rate for clarity of the exposition and $P(T)$ and $S(T)$ are the discount function and the survival function, respectively. By definition of the liquidity function, analogous to the discount function in the interest rate model, we have

$$l(n) = L(n+1)/L(n) \quad (29)$$

Then the stochastic movements of the liquidity factor are given by:

$$l(n, i) = l(n)(2/(1 + \delta^n))\delta^i \quad (30)$$

The valuation of liquidity contingent claims is given by

$$C(n - 1, i) = 0.5 l(n, i)(C(n, i + 1) + C(n, i)) \text{ for } i = 0, \dots, n - 1 \quad (31)$$

with the terminal condition for a call option, as an illustration

$$C(T, i) = \max(l(T, i) - X, 0) \text{ for } i = 0, 1, 2, \dots, T \quad (32)$$

Recent research shows that liquidity risk is an important factor in explaining the changes in the yield spread of corporate bonds. For example, Chen et al. (2007) empirically show that the liquidity risk is a significant factor to determine the yield spread variations of corporate bonds using more than 4000 corporate bonds. Their sample consists of both the investment grade and high-yield sectors. Covitz and Downing (2007) show that liquidity is important in the yield spread even for the commercial paper of less than 35 days to maturity. There should be a correlation between the liquidity risk and the credit risk since lower credit often relates to lower liquidity, as noted before. Modeling the joint stochastic movements will be discussed in the following section.

64.3.4 Energy Valuation Model

There are many energy markets with a range of forward contracts and option contracts. For example, there are the natural gas, oil, and electricity markets. See Eydeland and Wolyniec (2003) for an extensive discussion of the energy spot and derivative markets. To illustrate the application of the unified model to this broad financial sector, we will discuss only the electricity contracts here. Electricity is not a storable commodity, which has to be consumed when produced. The electricity futures are priced based on the volume bought over a time period at a delivery date. In comparing the forward price and the spot price, the “cost of carry” can be positive or negative depending on the month of the year, and therefore the forward prices can be higher or lower than the spot price. Electricity is a good example to use because it gives the contrast to the contingent claims pricing on financial instruments. Since electricity is not storable, the “cost of carry” cannot be explained by any dynamic trading strategies using electricity spot prices. However, for financial instruments, continuous arbitrage arguments can be used.

In applying the unified model, we assume that the electricity price is based on a fixed volume over one binomial period. The flow risk driver is the proportion of the electricity price at time n state i to the initial electricity price, denoted by $e(n, i)$, called the normalized forward price. Then the stochastic movements of the electricity price factor are given by:

$$e(n, i) = e(n)(2/(1 + \delta^n))\delta^i \quad (33)$$

The valuation of interest rate contingent claims is given by

$$C(n - 1, i) = 0.5 e(n, i)(C(n, i + 1) + C(n, i)) \text{ for } i = 0, \dots, n - 1 \quad (34)$$

with the terminal condition for a call option, as an illustration

$$C(T, i) = \max(e(T, i) - X, 0) \text{ for } i = 0, 1, 2, \dots, T \quad (35)$$

Electricity options on forward contracts can be used to calibrate the electricity price volatility, specified by the parameter δ . The importance of the energy model is to show that the Ho-Lee model does not apply only to the “rate risk” but to any flow risks, such as the cost of energy per unit time.

Energy contingent claims are growing in the market. With the gradual deregulation of the energy markets, where the energy prices are increasingly determined by the market, the use of forward contracts and energy options is growing. For example, a utility company may buy gas to generate electricity. However, this conversion depends on the prevailing relative prices of gas to electricity. Since electricity is not storable, often forward contracts and options on energy prices are used to ensure the availability of electricity at a certain period under stochastic spot prices and demands. Prices of gas vary stochastically depending on the suppliers. Utility companies may seek to buy the cheapest electricity among the suppliers. Contingent claims are used to provide options to the utility companies to seek the cheapest sources of gas.

64.3.5 Inflation Contingent Claims Valuation Model

The unified model can be applied to inflation contingent claims valuation. Let the inflation discount index be $\eta(n, i)$ a one binomial period at time n and state i . Let the forward inflation discount factor be $\eta(n)$. Then the stochastic movements of the inflation discount index are given by:

$$\eta(n, i) = \eta(n)(2/(1 + \delta^n))\delta^i \quad (36)$$

The valuation of inflation rate contingent claims is given by.

$$C(n - 1, i) = 0.5 \eta(n, i)(C(n, i + 1) + C(n, i)) \text{ for } i = 0, \dots, n - 1 \quad (37)$$

with the boundary condition.

$$C(T, i) = \max(\eta(T, i) - X, 0) \text{ for } i = 0, 1, 2, \dots, T \quad (38)$$

However, similar to the credit valuation model, the forward inflation index and the option of the inflation index are not directly observable in the capital markets. We need to relate the forward inflation index to tradable financial instruments.

Treasury Inflation Protection Securities (TIPS) are US Treasury bonds that have the principal adjusted to the inflation rates. For clarity of exposition, we make several simplifying assumptions here on the specification of TIPS. We assume that the bond has a coupon rate k for each binomial period, with an initial principal and maturity of \$1 and T periods, respectively. Furthermore, we assume that the convexity effect of TIPS is negligible such that a static inflation model can be used.

Then the value of TIPS is given by.

$$V(T) = c(1/\eta(1)P(1) + 1/(\eta(1)\eta(2))P(2) + \cdots + 1/(\eta(1)\cdots\eta(T))P(T)) + 1/(\eta(1)\cdots\eta(T))P(T) \quad (39)$$

where $P(n)$ is the discount function. Given the values $V(n)$ for $n = 0, 1, 2, \dots$, we can determine the $\eta(n)$.

The implied term structure of inflation rates can be estimated from the TIPS as many TIPS have been issued since their introduction in 1997. DePrince Jr. (2003) derives the term structure of inflation rates using TIPS from February 1997 to September 2002 and the paper shows that the implied inflation has economic content. Since inflation rate forecasts using surveys are only short term, the implied inflation rates can be useful for market participants. There are many contingent claims on inflation rates. Many insurance and pension products are indexed on inflation rates. Therefore a unified model incorporating inflation rates has many applications.

64.4 Modeling the Stochastic Process of the Unified Model

The main extension of the Ho-Lee model from the Black-Scholes model is based on the definition of the “underlying securities.” The Ho-Lee model does not require the underlying risk driver to be represented by tradable security. Instead, the Ho-Lee model determines the stochastic movements of the forward contracts on the flow risk driver based on the market observable forward prices of the risk factor.

Based on the risk-neutral measure, security drifts at a risk-free rate in the Black-Scholes framework. However, this is not a requirement for any risk factors in the Ho-Lee framework. For this reason, much of the interest rate model research focuses on modeling the stochastic movements of the risk driver when the Ho-Lee framework is used. We summarize some of the results in modeling the risk driver movements in this section.

The Ho-Lee (1986) unified model can be extended to a generalized model (Ho and Lee 2004b, 2007) that allows the risk driver to exhibit a broad range of behavior. We summarize some of the characteristics below. The model is presented in [Appendix B](#).

1. *State-time dependent volatilities.* In section B, we use the Ho-Lee model (1986) to illustrate the unified model. To capture a broader range of stochastic movements of the risk factor, δ in Eq. (9) does not have to be constant. It can be state and time-dependent, $\delta = \delta(n, i)$. Therefore, the model can be calibrated to an implied volatility function. In this case, the stochastic risk index can be shown to have the following functional form below

$$u(n, i) = u(n)F(n, i)\delta(n, i) \quad (40)$$

The function $F(n, i)$ is determined by the specification of the $\delta(n, i)$. An implied volatility function that declines with time n can induce a mean reversion behavior of the risk factor. This observation is made in the Black, Derman, and Toy model (1990).

2. *Regime switching:* The model avoids explosively high rates exhibited by the lognormal models on the one hand and negative rates exhibited by the normal models on the other. Since we can construct δ to be dependent on the state, the forward volatilities of each factor can be proportional to the rate level. When the rates are low, the process exhibits a lognormal behavior. When the rates are high, the process exhibits a normal distribution behavior. Cheyette (1997) shows that interest rate movements tend to change the stochastic process regime depending on the interest rate level. Ho and Mudavanhu (2007) show empirically that the specification of the change in regime in the interest rate distribution can be implied from the swaption markets.
3. *Multi-movement model:* The movements of a risk factor can have multiple risk drivers. At each node, each factor can take a binary movement. Let e be a movement over one step, represented by 0 or 1; and there are four possible states emulating from each node for a two-factor model represented by $v = (e_1, e_2)$. Let e^* represent the opposite state of e such that $e^* = 0$ and 1 when $e = 1$ and 0, respectively. Then the two consecutive movements of (e_1, e_2) followed by (e^*_1, e^*_2) must meet at the same node. This is the recombining property that reduces the possible future scenarios for analysis significantly without a comparable loss of accuracy of the model.

For clarity of exposition, we have presented the two-factor model which is $u(n, i, j) = u(n)F(n, i)F(n, j)\delta(n, i)\delta(n, j)$. The generalization of the two-factor model to the multifactor model for Eq. (9) is straightforward. For an m -factor model, (9) is represented by

$$u(n, i_1, \dots, i_m) = u(n)C(n, m)\Lambda(i_1, \dots, i_m) \quad (41)$$

where $C(n, m) = \left[\frac{2}{1+\delta_1^n} \right] \dots \left[\frac{2}{1+\delta_m^n} \right]$ and $\Lambda(i_1, \dots, i_m) = \delta_1^{i_1} \dots \delta_m^{i_m}$. This shows that $u(n, i_1, \dots, i_m)$ has three components. First, $u(n)$ represents the forward price based on the initial term structure. Second, $C(n, m)$ is the convexity term given by $\left[\frac{2}{1+\delta_1^n} \right] \dots \left[\frac{2}{1+\delta_m^n} \right]$. This term adjusts the price such that the additional returns derived

from the convexity are countered by this adjustment to assure the arbitrage-free condition. Third, the stochastic movement term $\Lambda(i_1, \dots, i_m)$ is given by $\delta_1^{i_1} \dots \delta_m^{i_m}$. The arbitrage-free condition for the model in Eq. (41) is proved in Ho-Lee (2007). The model for multiple risk drivers can be defined analogously. For the term structure of interest rates, Litterman and Scheinkman (1991) have shown that historically the yield curve exhibits three principal movements. Ho and Mudavanu (2007) describe a multi-movement model that is consistent with this empirical observation where the arbitrage-free yield curve movements can be parallel movements or steepening movements, and the movements are implied from the swaption market using multiple implied volatility functions. *Multifactor model*: We use the analogous multidimensional recombining lattice described in subsection 3. Combining multiple risk factors is also relatively straightforward. Let the risk factor be combined of m risk factors, $u^1(n, i) \dots u^m(n, i_m)$, then the model is:

$$u(n, i_1, \dots, i_m) = u^{i_1}(n) \dots u^{i_m}(n) F^{i_1}(n, i_1) \dots F^{i_m}(n, i_m) \delta^{i_1}(n, i_1) \dots \delta^{i_m}(n, i_m) \quad (42)$$

Therefore the unified model enables us to value securities subjected to several types of risks. Ho and Lee (2007) provide a more detailed description of the multifactor model. When the risk factors are uncorrelated, then all the binomial probabilities are the same. Otherwise, appropriate binomial probabilities give the desired correlations (see Ho and Lee 2004a, Note that when there are correlations between two factors, then non-option embedded security in one factor may have embedded options from another factor. For example, when there is a correlation between credit risk and interest rate risk, a CDS may have embedded interest rate options. The option is induced by the recovery payments at default viewed as interest rate risk “prepayment option” when the default rate changes with the interest rate level. However, such option value is typically very small. Also, this observation does not mean that the methodology presented is inappropriate. We just need to take this observation into account in the calibration process.

4. *Arbitrage-free condition*. The model takes the forward discount factors and their associated term structure volatilities as inputs to the valuation. The projected discount rate of a fixed-income instrument is given as the combination of the stochastic shifts from the sum of the forward curves of the factors with an adjustment for the convexity effect to ensure the arbitrage-free condition is held. The value derived from the roll-back valuation algorithm using the lattice model is always consistent with these input parameters.
5. *Term structure of rates and the term structure movements*: In Sect. 64.3, we have focused on the stochastic movements of the discount indices by specifying the risk driver at each binomial node (n, i) . Using the rollback process, the entire term structure of the discount index can be determined at each node. For example, for the interest rate model, the term structure of interest rate can be determined at each node. For the credit model, the credit derivative swap curve can also be specified at each node of the binomial lattice. Therefore, the contingent claims can be specified for the entire term structure and not confined only to the one-period risk

driver. Furthermore, for a multifactor model, at each node $(i, j, k \cdot \cdot)$, multiple movements of the term structures can be modeled.

64.5 Implications of the Unified Model to Risk Management

64.5.1 Relevance of the Unified Model to the Financial Crisis 2008

A tranche of a CDO has an embedded option within a CDO package, which in turn is pooled from asset-backed security (ABS) tranches that themselves have embedded options. Each ABS is a pool of mortgage loans or bonds, which have also embedded options. These options are driven by multiple risk drivers. For this reason, the actuarial credit risk measure would not be inappropriate to identify the risks of these securities that led to the subprime crisis.

Managing credit, market, and liquidity risks separately would not be able to provide a coherent valuation and analytical framework. For example, Basel II requirements suggest entirely different risk measurement approaches to market risk, credit risk, and liquidity risks, and the Basel requirements provide no guidance in rolling up the risks to the enterprise level. As a result, financial institutions have separate market risk and credit risk departments. Such separation of risk analyses can easily fail to properly identify the risk of structured products such as asset-backed securities and collateralized debt obligations which have complicated embedded credit and interest rate options.

A coherent way of isolating the credit risk from interest rate risk is important. For example, an internal measure of the credit guarantee cost is a useful way to bifurcate the credit option value from the interest rate option in a bond. Federal Home Loan Bank (FHLB) Mortgage Partnership Finance (MPF) lending programs to their network of banks uses this approach. This approach allows their affiliated banks to retain the credit risk while selling the interest rate options to the FHLBs. The unified model can be used to identify these values.

The application of the unified model goes beyond the valuation of securities. Enterprises can be modeled as contingent claims of multiple flow risk drivers. For a financial institution, often the economic capital is defined as the asset net of the liability. But in many instances, the liability cannot be valued separately from the assets. For example, the run on a bank behavior of the depositors can be viewed as the depositors' American put option on the banks' asset value with a strike price of the account value. The put option adds additional value to the deposit liability and lowering the appropriate measure of the economic capital. Another example is the margin requirements on some liabilities. Often, financial institutions' funding requires margins, which increase with the value of the assets net of the face value of the liability. As a result of this requirement, the financial institution has to sell assets to generate cash as the asset liquidity spread widens. This process is called "deleveraging" which has shown to be very significant in the 2008 financial crisis. The descriptions of these implications of these embedded options in the economic capital are beyond the scope of this Sect. 64.3.

But we have highlighted the importance of using contingent claims analysis to the analysis on economic capital and risk management of a financial institution.

Financial research has studied corporate liabilities in terms of the Merton model, where equity is viewed as a call option on the firm value, the total capitalization, with the strike price equaling the face value of a bond. As the 2008 financial crisis shows, distressed financial institutions have value based more on their economic capital value than on the total capitalization. As noted above, the economic capital values are contingent claims on multiple flow risk drivers, where a unified model may be more appropriate for analysis than the Merton model that is based on the stock risk drivers.

64.5.2 Benchmark Securities and Model Calibration

A portfolio of these fixed-income securities with similar liquidity risk and credit risk that have market prices is called a benchmark portfolio. The term structures of volatilities of interest rates, default rates, and liquidity rates can be calibrated from the prices of the fixed-income securities in the benchmark portfolio. The calibrated model can then be used to value other financial instruments that have similar credit risk and liquidity risk. Any discrepancy between the model value and the market price is called the “cheap/rich” value. Therefore, the financial instrument is relatively valued to the benchmark portfolio, and that the model determines the risk-neutral measures for the valuation purpose.

The unified model is calibrated to the appropriate benchmark portfolio that is chosen to solve the problem at hand. This approach recognizes that there are often other factors affecting the quoted prices, such as the bid-ask spreads, commissions, and other microstructure factors. The benchmark portfolio should contain securities relevant to the instruments to be relatively valued. For example, CDSs are not all liquid. The choice of CDSs used in the benchmark portfolio should take liquidity into account. In this sense, the unified model should be viewed as a generalized relative valuation model. The model provides the value of a security relative to a benchmark portfolio.

The calibrated term structure of default rates and liquidity rates should be upward sloping in normal market conditions. This is because the default risk and liquidity cost both increase with time to maturity. As a result, the forward default rates and liquidity rates must also increase in calendar time projected into the future, as the maturity of the fixed-income security shortens. But when a fixed-income security maturity shortens, we expect that both the credit spread and the liquidity spread tighten, not rising as predicted by the model under risk-neutral measure, as opposed to the market measure. By the same token, the normal yield curve is upward sloping. The arbitrage-free interest rate models would consistently “predict” that the short-term rates would rise in the future, and clearly, this prediction is inconsistent with the historical experience. This apparent contradiction does not affect the robustness of the model which is based on the risk-neutral measure.

64.5.3 Combining Interest Rate Risk, Credit Risk, and Liquidity Risk in Enterprise Risk Management

Historically, from the deregulation of interest rates in the late 1970s till the mid-1990s, the USA had experienced significant interest rate risks fluctuating between 6% and 12%. However, in the 2000s, interest rates have fallen to less than 4% in recent months. At this low-interest rate regime, credit risk and liquidity risk have become significant contributors to the value of many financial instruments. That is, credit risk and liquidity risk factors have to be considered along with interest rate factors. The Ho-Lee framework is therefore proposed here to deal with this current market condition.

Let us consider the valuation of a callable bond to illustrate the economics of the unified model. Let the bond has a maturity of N years and is callable at par any time after M year, with $M < N$. The bond has a coupon rate such that the bond has a value at par. The bond is subject to interest rate, credit, and liquidity risks.

The bond has an embedded call option. Interest rate models would consider the options subject only to the interest rate risk. The term structure of volatilities is estimated from the swaption market prices. However, the bond would be called if the credit risk premium or the liquidity premium has been tightened. The unified model would take the combined effect into considerations.

Valuation of fixed-income securities, for example, corporate bonds with a call or put options, can be determined by the standard roll-back procedure using the lattice model. The arbitrage-free condition would ensure that when the bond's call option is stripped off, the bond value calculated by the roll-back method is identical to the standard cash flow bond model. Therefore, all arbitrage-free conditions would apply. For example, the put-call parity of options is automatically held. Since the arbitrage-free condition is satisfied at each node point, the roll-back method using Bellman's optimization can be used.

Thus far, we have described the valuation on a contingent claim on the term structures of interest rate, default rates, and liquidity rates. We now extend this discussion to describe the valuation of a portfolio of credit and liquidity risks. Arbitrage-free interest rate models can extend from a single security valuation to that for a portfolio simply because the identical yield curve applies to all the securities. Such is not the case with credit and liquidity risks. The term structure of default rates and liquidity rates are not the same, not even perfectly correlated, across securities. A model of a portfolio of risk factors has to be described.

For simplicity, let us consider a portfolio of two bonds. To determine VaR, we assume that the risk sources are the discount function, the survival function, and the liquidity function. Consider the log of these three functions. We assume that they are uniformly distributed and are constant in time. The associated term structures of volatility are flat for simplicity.

To determine the individual VaR, such as market, credit, and liquidity VaR, we perturb each risk source while the other two risk sources are kept constant. To

determine the interest rate VaR, we keep the other survival and liquidity functions fixed. Since the discount function is uniformly distributed, we generate the profit/loss distribution from the discount function disturbances.⁵ Once we generate the distribution of the profit/loss, we can determine the maximum possible loss given the significance level, which is the Market VaR. Similarly, we can get credit VaR and liquidity VaR of an individual bond.

Since the three risk sources determine the bond value and the discount function is applicable across the bonds, we have altogether five random variables, which are uniformly distributed by assumption. We use the Gaussian Copula method to generate the joint distribution of five random variables. Specifically, we map the uniform distribution into the normal distribution on a percentile-to-percentile basis. Then, we generate multivariate normal distribution with correlation coefficients being historical correlations for simplicity. We determine the portfolio VaR from the joint cumulative distribution, given the significance level. The diversification effect is the difference between the portfolio VaR and the sum of individual VaR. The diversification effect depends on the correlation structure among the risk sources.

The component VaR is defined by the contribution part of the portfolio VaR from each risk source. To determine the component VaR from a risk source, we eliminate the risk source while we keep the other risk sources. The decrease in the portfolio VaR can be regarded as the contribution of the risk source to the portfolio VaR. Then we eliminate another risk source to determine the component VaR of the second risk source. Repeating the same procedure until we eliminate the last risk source, we can determine the component VaR of each risk source.

64.5.4 Financial Engineering Considerations

A unified model must necessarily require extensive computational resources. And therefore, a specification of the unified model must be related to issues in financial engineering, as well as the principles in financial theories. Note that the unified model is constructed based on binomial periods but the model does not require the binomial model to have a constant calendar time period. For example, the binomial steps do not have to be one month for all periods. They can vary and therefore the unified model can be adapted to a variable step size construction quite simply.

Consider the callable bond in this illustration. Since the call protection period is only one year, the model must measure this one-year short-dated option accurately. The unified model with variable step size would use a small lattice step size in the early years. For the later year, longer step size is used to save the computing time without consequential loss of accuracy.

Since a recombining lattice enumerates all possible risk scenarios with the assigned risk-neutral measures, within the context of the model, we can assign probabilities to each risk scenario. Furthermore, based on any measure of

⁵The three-factor unified Ho and Lee model is given in [Appendix B](#).

distance between two risk scenarios, we can form equivalent classes in this risk space with their representative risk scenarios. See Ho (1992) for such a methodology. As a result, we can determine the set of representative scenarios with their assigned probability weights for valuation when scenario paths have to be used, as in the case of valuation a mortgage loan using a prepayment model. The structured sampling approach can enhance the computational efficiency significantly.

64.6 Conclusions

This chapter shows that risk drivers can be dichotomized into “stock” or “flow” attributes. The Black-Scholes model is a relative valuation model for the stock risk drivers and the Ho-Lee model for the flow risk drivers. A unified model that values contingent claims on multiple flow risk drivers is presented here.

The model is then applied to five different flow risk drivers: interest rate, credit risk, liquidity risk, energy risk, and inflation risk. We show how the unified model can be adjusted for each type of flow risk driver and how the benchmark securities are used to calibrate the models. Furthermore, we show that the unified model provides the valuation of contingent claims on multiple flow risk drivers. And hence the unified model provides a model and a framework to analyze the combined effect of multiple flow risk drivers on securities and the economic capital of a financial institution.

The 2008 financial crisis illustrates the importance of the applications of the unified model. For example, CDO and ABS are structured products that have significant credit and interest rate embedded options. This example shows that risk management should not take a silo approach to manage market risk, credit risk, and liquidity risk in isolation. This is because the impact of the sum of these risks is not the same as that of the combined risks. The options embedded in the security do not offer a simple separation of the effects of the risk drivers.

Furthermore, we show that the economic capital of a financial institution is a contingent claim on multiple risk drivers and therefore the economic capital cannot be measured simply as the asset net of the liability. The unified model should be used to analyze a distressed financial institution along with the Merton model which is a relative valuation model on the stock risk driver and not modeling the economic capital as a contingent claim on flow risk drivers. The unified model can provide a coherent framework to manage enterprise risk. In particular, we show how the model is used to generate the value-at-risk measures for the combined market, credit, and liquidity risks.

Appendix A: A General Form of the Arrow-Debreu Primitive Securities

The Arrow-Debreu primitive securities at time n are equal to a vector with $(n + 1)$ elements, which we can calculate by multiplying the following n matrixes.

$$\begin{aligned}
& \left[\frac{e^{-\rho(0)}}{1+\delta^0} \quad \frac{e^{-\rho(0)}}{1+\delta^0} \right]_{1 \times 2} \times \left[\begin{array}{cc|c} \frac{e^{-\rho(1)}\delta}{1+\delta^1} & \frac{e^{-\rho(1)}\delta}{1+\delta^1} & 0 \\ 0 & \frac{e^{-\rho(1)}}{1+\delta^1} & \frac{e^{-\rho(1)}}{1+\delta^1} \end{array} \right]_{2 \times 3} \times \left[\begin{array}{cc|cc} \frac{e^{-\rho(2)}\delta^2}{1+\delta^2} & \frac{e^{-\rho(2)}\delta^2}{1+\delta^2} & 0 & 0 \\ 0 & \frac{e^{-\rho(2)}\delta}{1+\delta^2} & \frac{e^{-\rho(2)}\delta}{1+\delta^2} & 0 \\ 0 & 0 & \frac{e^{-\rho(2)}}{1+\delta^2} & \frac{e^{-\rho(2)}}{1+\delta^2} \end{array} \right]_{3 \times 4} \\
& \times \left[\begin{array}{ccc|cc} \frac{e^{-\rho(3)}\delta^3}{1+\delta^3} & \frac{e^{-\rho(3)}\delta^3}{1+\delta^3} & 0 & 0 & 0 \\ 0 & \frac{e^{-\rho(3)}\delta^2}{1+\delta^3} & \frac{e^{-\rho(3)}\delta^2}{1+\delta^3} & 0 & 0 \\ 0 & 0 & \frac{e^{-\rho(3)}\delta^1}{1+\delta^3} & \frac{e^{-\rho(3)}\delta^1}{1+\delta^3} & 0 \\ 0 & 0 & 0 & \frac{e^{-\rho(3)}}{1+\delta^3} & \frac{e^{-\rho(3)}}{1+\delta^3} \end{array} \right]_{4 \times 5} \dots\dots\dots \\
& \dots\dots\dots \times \left[\begin{array}{cccc|cc} \frac{e^{-\rho(n-1)}\delta^{n-1}}{1+\delta^{n-1}} & \frac{e^{-\rho(n-1)}\delta^{n-1}}{1+\delta^{n-1}} & 0 & \dots & 0 & 0 \\ 0 & \frac{e^{-\rho(n-1)}\delta^{n-2}}{1+\delta^{n-1}} & \frac{e^{-\rho(n-1)}\delta^{n-1}}{1+\delta^{n-1}} & \dots & 0 & 0 \\ \vdots & \vdots & \frac{e^{-\rho(n-1)}\delta^{n-3}}{1+\delta^{n-1}} & \frac{e^{-\rho(n-1)}\delta^{n-3}}{1+\delta^{n-1}} & \vdots & \vdots \\ 0 & 0 & \dots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & \frac{e^{-\rho(n-1)}}{1+\delta^{n-1}} & \frac{e^{-\rho(n-1)}}{1+\delta^{n-1}} \end{array} \right]_{n \times (n+1)}
\end{aligned}$$

For example, when n is equal to 1, the Arrow-Debreu primitive securities at time 1 is the first matrix such that $A(1, 1)$ is equal to $\frac{e^{-\rho(0)}}{1+\delta^\sigma}$ and $A(1, 0)$ is equal to $\frac{e^{-\rho(0)}}{1+\delta^\sigma}$. When n is equal to 2, we can generate the Arrow-Debreu primitive securities at time 2 by multiplying the first two matrixes such that

$$\begin{aligned}
& \begin{bmatrix} \frac{e^{-\rho(0)}}{1+\delta^0} & \frac{e^{-\rho(0)}}{1+\delta^0} \end{bmatrix} \times \begin{bmatrix} \frac{e^{-\rho(1)}\delta}{1+\delta^1} & \frac{e^{-\rho(1)}\delta}{1+\delta^1} & 0 \\ 0 & \frac{e^{-\rho(1)}}{1+\delta^1} & \frac{e^{-\rho(1)}}{1+\delta^1} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sum_{t=0}^1 -\rho(t)}{e^{\sum_{t=0}^1 -\rho(t)}} \delta & \frac{\sum_{t=0}^1 -\rho(t)}{e^{\sum_{t=0}^1 -\rho(t)}} (1+\delta) & \frac{\sum_{t=0}^1 -\rho(t)}{e^{\sum_{t=0}^1 -\rho(t)}} 1 \\ \prod_{t=1}^1 1+\delta^t & \prod_{t=1}^1 1+\delta^t & \prod_{t=1}^1 1+\delta^t \end{bmatrix} \\
& \qquad \qquad \qquad 1 \times 2 \qquad \qquad \qquad 2 \times 3 \qquad \qquad \qquad 1 \times 3
\end{aligned}$$

The first element of the vector is $A(2, 2)$, the second element is $A(2, 1)$, and the last element is $A(2, 0)$. Similarly, we can generate the Arrow-Debreu primitive securities at the following periods and determine them accordingly. The intuition behind the general form of the Arrow-Debreu primitive securities is the recursive relationship among $A(n, i)$, $A(n-1, i)$, and $A(n-1, i-1)$ for any n and i such as $A(n, i) = \frac{e^{-r(n-1)}\delta^i}{(1+\delta^{n-1})}A(n-1, i) + \frac{e^{-r(n-1)}\delta^{i-1}}{(1+\delta^{n-1})}A(n-1, i-1)$.

Appendix B: Three-Factor Unified Model

The model is provided in Ho-Lee (2007). For completeness, we specify the model in this appendix. Let $B_{i,j,l}^n(T)$ be the T year bond price at time n , at state (i, j, k) . Then $B_{i,j,l}^n(T)$ is specified by combining three one-factor models. Specifically, we have

$$\begin{aligned}
B_{i,j,l}^n(T) &= \frac{P(n+T)}{P(n)} \prod_{k=1}^n \left(\frac{1+\delta_{0,r}^{k-1}(n-k)}{1+\delta_{0,r}^{k-1}(n-k+T)} \right) \prod_{k=0}^{i-1} \delta_{k,r}^{n-1}(T) \\
&\times \frac{S(n-1+T)}{S(n-1)} \prod_{k=1}^n \left(\frac{1+\delta_{0,h}^{k-1}(n-k)}{1+\delta_{0,h}^{k-1}(n-k+T)} \right) \prod_{k=0}^{j-1} \delta_{k,h}^{n-1}(T) \\
&\times \frac{L(n+T)}{L(n)} \prod_{k=1}^n \left(\frac{1+\delta_{0,l}^{k-1}(n-k)}{1+\delta_{0,l}^{k-1}(n-k+T)} \right) \prod_{k=0}^{l-1} \delta_{k,l}^{n-1}(T)
\end{aligned} \tag{43}$$

where

$$\begin{aligned}
\delta_{i,r}^n(T) &= \delta_{i,r}^n \delta_{i,r}^{n+1}(T-1) \left(\frac{1+\delta_{i+1,r}^{n+1}(T-1)}{1+\delta_{i,r}^{n+1}(T-1)} \right) \\
\delta_{i,h}^n(T) &= \delta_{i,h}^n \delta_{i,h}^{n+1}(T-1) \left(\frac{1+\delta_{i+1,h}^{n+1}(T-1)}{1+\delta_{i,h}^{n+1}(T-1)} \right) \\
\delta_{i,l}^n(T) &= \delta_{i,l}^n \delta_{i,l}^{n+1}(T-1) \left(\frac{1+\delta_{i+1,l}^{n+1}(T-1)}{1+\delta_{i,l}^{n+1}(T-1)} \right)
\end{aligned} \tag{44}$$

and the one-period forward volatilities are given by definition,

$$\begin{aligned}\delta_{i,r}^m(1) &= \delta_{i,r}^m = \exp\left(-2 \cdot \sigma_r(m) \min\left(h_{i,r}^m, H_r\right) \Delta t^{3/2}\right) \\ \delta_{i,h}^m(1) &= \delta_{i,h}^m = \exp\left(-2 \cdot \sigma_h(m) \min\left(h_{i,h}^m, H_h\right) \Delta t^{3/2}\right) \\ \delta_{i,l}^m(1) &= \delta_{i,l}^m = \exp\left(-2 \cdot \sigma_l(m) \min\left(h_{i,l}^m, H_l\right) \Delta t^{3/2}\right)\end{aligned}\quad (45)$$

Where the functions $\sigma_m(n) = (a + bn) \exp(-cn) + d$ where $m = r, h, l$ is specified by the parameters a, b, c , and d , which can be obtained from the calibration to the market price of an option on CDS, etc. This specification of the implied volatility function allows for a broad range of shapes including downward sloping or dumped shapes.

Using the direct extension, we can specify the one-period hazard rates for the two-factor model for any future period m and state i , and $h_{i,1}^m$ and $h_{i,2}^m$ are defined by

$$\begin{aligned}h_{i,r}^m \Delta t &= -\log\left(\frac{P(m+1)}{P(m)}\right) - \sum_{k=1}^m \log\left(\frac{1 + \delta_{0,r}^{k-1}(m-k)}{1 + \delta_{0,r}^{k-1}(m-k+1)}\right) - \sum_{k=0}^{i-1} \log\left(\delta_{k,r}^{m-1}(1)\right) \\ h_{i,h}^m \Delta t &= -\log\left(\frac{S(m)}{S(m-1)}\right) - \sum_{k=1}^m \log\left(\frac{1 + \delta_{0,h}^{k-1}(m-k)}{1 + \delta_{0,h}^{k-1}(m-k+1)}\right) - \sum_{k=0}^{i-1} \log\left(\delta_{k,h}^{m-1}(1)\right) \\ h_{i,l}^m \Delta t &= -\log\left(\frac{L(m)}{L(m-1)}\right) - \sum_{k=1}^m \log\left(\frac{1 + \delta_{0,l}^{k-1}(m-k)}{1 + \delta_{0,l}^{k-1}(m-k+1)}\right) - \sum_{k=0}^{i-1} \log\left(\delta_{k,l}^{m-1}(1)\right)\end{aligned}\quad (46)$$

References

- Black, F., E. Derman, and W. Toy. 1990. A one-factor model of interest rates and its application to Treasury bond options. *Financial Analysts Journal* 46: 33–39.
- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–654.
- Brace, A., D. Gatarek, and M. Musiela. 1997. The market model of interest rate dynamics. *Mathematical Finance* 7: 127–155.
- Chen, L., D. Lesmond, and J. Wei. 2007. Corporate yield spreads and bond liquidity. *Journal of Finance* 62: 119–149.
- Cheyette, O. 1997. Interest rate models. In *Advances in fixed Income valuation, modeling, and risk management*, ed. Frank J. Fabozzi. New Hope: Frank J. Fabozzi Associates.
- Covitz, D., and C. Downing. 2007. Liquidity or credit risk? The determinant of very short-term corporate yield spreads. *Journal of Finance* 62: 2303–2328.
- Cox, J., J. Ingersoll, and S. Ross. 1985. A theory of term structure of interest rates. *Econometrica* 53: 363–384.
- Das, S., L. Freed, G. Geng, and N. Kapadia. 2006. Correlated default risk. *The Journal of Fixed Income* 16: 7–32.
- DePrince, A., Jr. 2003. Assessing the term structure of expected inflation using Treasury inflation-protected securities. *Business Economics* 38: 46–54.

- Duffie, D. 2005. Credit risk modeling with affine processes. *Journal of Banking and Finance* 29: 2751–2802.
- Eydeland, A., and K. Wolyniec. 2003. *Energy and power risk management: New developments in modeling, pricing, and hedging*. Hoboken: John Wiley & Sons, Inc.
- Harrison, J., and D. Kreps. 1979. Martingales and arbitrage in multi-period securities markets. *Journal of Economic Theory* 20: 381–408.
- Heath, D., R. Jarrow, and A. Morton. 1992. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica* 60: 77–105.
- Ho, T.S.Y. 1992. Managing Illiquid Bonds and Linear Path Space. *The Journal of Fixed Income* 2: 80–94.
- Ho, T.S.Y., and S.B. Lee. 1986. Term structure movements and pricing interest rate contingent claims. *The Journal of Finance* 41: 1011–1029.
- . 2004a. *The Oxford guide to financial modeling*. New York: Oxford University Press.
- . 2004b. A closed-form multifactor binomial interest rate model. *The Journal of Fixed Income* 14: 8–16.
- . 2007. Generalized Ho-Lee model: A multi-factor state-time dependent implied volatility function approach. *The Journal of Fixed Income* 17: 18–37.
- . 2009a. A unified credit and interest rate arbitrage-free contingent claim model. *The Journal of Fixed Income* 18: 5–17.
- . 2009b. Valuation of credit contingent claims: An arbitrage-free credit model. *Journal of Investment Management* 7: 49–65.
- Ho, T.S.Y., and B. Mudavanu. 2007. Interest rate models' implied volatility function stochastic movements. *Journal of Investment Management* 5: 1–22.
- Litterman, R., and J.A. Scheinkman. 1991. Common factors affecting bond returns. *The Journal of Fixed Income* 1: 54–61.
- Longstaff, F., and A. Rajan. 2008. An empirical analysis of the pricing of collateralized debt obligation. *Journal of Finance* 63: 529–563.
- Rebonato, R. 1998. *Interest rate option models: Understanding, analyzing and using models for exotic interest-rate options*. Chichester: John Wiley.
- Vasicek, O. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5: 177–188.



A Comparison of Formulas to Compute Implied Standard Deviation

65

James S. Ang, Gwoduan David Jou, and Tsong-Yue Lai

Abstract

We derive an exact closed-form solution for the implied standard deviation in the Black and Scholes' option pricing model under the condition that the underlying asset price equals the present value of the exercise price. The exact closed-form solution provides the true implied standard deviation and has no estimation error. We then relax this condition and develop three new formulas that depend on a Taylor series expansion utilizing one, two or three options. Simulations show these formulas produce lower estimation errors than extant approaches, and the third formula gives the best overall results under different parameter values.

Keywords

Implied standard deviation · Implied volatility · Options · Option pricing model · Taylor formula

We are grateful for the helpful comments and suggestions by John Erickson, Shean-Bii Chiu, Shyan-yuan Lee, Tsun-Siou Lee, Po-Hsun Lin and Sharon Lai. We are indebted to Chih-Yuan Wu for excellent research assistance. This research was supported by National Science Council (contract number NSC89-2811-H-002-0006) in Taiwan.

J. S. Ang (✉)

Department of Finance, College of Business, Florida State University, Tallahassee, FL, USA
e-mail: Jang@cob.fsu.edu

G. D. Jou

CFO, Taikang Life Insurance Co., Beijing, China
e-mail: Jou@MBA.ntu.edu.tw

T.-Y. Lai

Department of Finance, California State University, Fullerton, Fullerton, CA, USA
e-mail: Tylai@fullerton.edu

65.1 Introduction

Standard deviation of the underlying asset return plays two roles: to estimate the option pricing and to capture the volatility of the underlying asset in the spot market. In Merton (1973) and Black-Scholes (1973) option pricing model, the standard deviation of the underlying asset's rate of return is the only parameter, which cannot be observed directly. How to estimate an accurate standard deviation becomes a primary concern in option pricing and trading. Three approaches have been developed in the finance literature to estimate standard deviation: (1) the historical standard deviation, (2) the implied standard deviation (ISD) derived from equating the Black-Scholes' option pricing model with the market value of the call option, and (3) the multiple implied standard deviations from options with different exercise prices. Garman and Klass (1980) study the historical standard deviation by using open, high, low prices, and closed prices' data to estimate the standard deviation. To support the use of historical standard deviation for implied standard deviation in option pricing model requires that the underlying asset's rate of return is stationary over the option's life, which contradicts the time-varying standard deviation documented by Schwert (1989).

Implied standard deviation has been studied by Latane and Rendleman (1976), Beckers (1981), Manaster and Koehler (1982), Brenner and Subrahmanyam (1988), Lai et al. (1992), Chance (1996), Hallerback (2004), Corrado and Miller (1996, 2004), and Li (2005). Since the Black-Scholes' option pricing model is a nonlinear equation, an explicit analytic solution for the ISD is not available in the literature (except for at-the-money call) and numerical methods are used to approximate the ISD. By applying the Newton-Raphson method, Manaster and Koehler (1982) provide an iterative algorithm for the ISD. Brenner and Subrahmanyam (1988) applied Taylor series expansion at zero base to the cumulative normal function in pricing Black-Scholes option pricing model up to the first-order term and set the underlying asset price to equal the present value of exercise price to solve the ISD. Following the same approach as Brenner and Subrahmanyam, Li (2005) expands the expression to third order term and solve for the ISD with a cubic equation. The accuracy of ISD derived by Brenner and Subrahmanyam (1988) and Li (2005) depends not only on the deviation of stock price (S) from the present value (K) of exercise price (E), but also on the magnitude of the ISD. Since Li includes third order term in the Taylor expansion on the cumulative normal distribution in his derivation, Li claims that his formula of ISD provides a consistently more accurate estimate of the true ISD than that of Brenner and Subrahmanyam's formula. It is worthy to note that both formulas are approximate rather than exact closed-form solution for ISD. Therefore, there exists certain amount of estimation error by using either formulas to test the ISD. Lai et al. (1992) derive a closed form solution for the ISD in terms of the delta $\partial C/\partial S$ and $\partial C/\partial E$. Given $C = S(\partial C/\partial S) + E(\partial C/\partial E)$, their closed-form solution can be simplified to depend on either $\partial C/\partial S$ or $\partial C/\partial E$ only.

To allow for the deviation between the underlying asset price and the present value of exercise price, Corrado and Miller (1996) follow Brenner and Subrahmanyam's approach by expanding the cumulative normal function at zero to the first order term to derive a quadratic equation of the ISD. The ISD is then

obtained by solving the quadratic equation. However, to improve accuracy, Corrado and Miller made two substantial substitutions in their model.¹ First, $\ln(S/K)$ is replaced by $2(S - K)/(S + K)$. Second, $4[(S - K)/(S + K)]^2$ is replaced by $\alpha [(S - K)/(S + K)]^2$ that minimizes the concavity at $S = K$. The accuracy of their improved formula depends on the accuracy of these substitutions. Hallerback (2004) also derives an improved formula, which is similar to Corrado and Miller's formula (1996), to compute the ISD. Corrado and Miller (2004) argue that the formula derived by Hallerback (2004) collapse to their formula if (1) the geometric mean (i.e., $\sqrt{SK}$) of the stock price and the present value of exercise price K in Hallerback's formula is substituted by the arithmetic mean (i.e., $(S + K)/2$), and (2) the parameter value of 1.85 in Hallerback's formula is substituted by the tweaking parameter α in Corrado and Miller's (2004) formula. However, this argument is questionable because the arithmetic mean is never less than the geometric mean. Li (2005) follows Corrado and Miller's (1996) approach by expanding the Black-Scholes option pricing at the K/S to the third order. Li derives a formula of ISD which is slightly different from that of Corrado and Miller due to an extra term. However, Li acknowledges that the extra term in his formula could be ignored in most of the cases involving deep-in or out-of-the money options.²

Chance (1996) assumes different exercise prices result in different ISDs, which violates the constant variance assumption used in deriving the Black-Scholes' option pricing model. Under the existence of a call at-the-money assumption, Chance uses Brenner and Subrahmanyam's formula to calculate the at-the-money's ISD. Chance then applies Taylor series expansion to the difference of the call option (i.e., ΔC) in terms of the $\partial C/\partial S$, $\partial C/\partial E$, and the second order terms; all are calculated at-the-money. In order to apply Chance's formula³ to compute the ISD, the standard deviation and the option price at-the-money must be given. In other words, if the underlying asset price deviates from the present value of the exercise price and the call option price is not available (or unobservable) in the market, then Chance's formula for the ISD may not apply.

In this chapter,⁴ we initially derive an exact closed-form solution for the ISD under the condition that the underlying asset price equals the present value of the exercise price. We relax this restrictive condition, and derive three alternate formulas to estimate the ISD. Applying Taylor series expansion on the Black-Scholes' option pricing model derives the first formula to estimate the ISD. The difference between this formula and previously derived formulas is that we apply Taylor's formula to the cumulative normal functions $N[\ln(S/K)/(\sigma\sqrt{T}) + \sigma\sqrt{T}]$ and $N[\ln(S/K)/(\sigma\sqrt{T}) - \sigma\sqrt{T}]$ at base

¹In detail see the [Appendix](#).

²See Eq. (15) and the footnote 3 in Li (2005). Li provides two formulas (10) and (16) for ISD, one is at-the-money and one is nearly at-the-money. Both formulas (10) and (16) in his paper include a term of $\cos^{-1}[\cos^{-1}(\sqrt{\pi 3\alpha/32})/3]$.

³Chance uses $\sigma = C^* \sqrt{2\pi}/(S\sqrt{T})$ in his formula, where C^* is the price of the at-the-money call.

⁴This chapter extends the analysis in Ang et al. (2009).

$\ln(S/K)/(\sigma\sqrt{T})$ up to the second-order terms⁵ rather than at the zero point up to the first-order term. The deviation of this approach from the true cumulative normal value is of the magnitude $\pm\sigma\sqrt{T}$ rather than the larger $\ln(S/K)/(\sigma\sqrt{T}) \pm \sigma\sqrt{T}$. The inclusion of the second-order terms can partially capture the non-linearity of the cumulative normal function and should improve the accuracy of the ISD. In sum, the advantages of this approach are to circumvent the ad hoc replacements present in CM and to improve the accuracy of the ISD's estimation.

Under the Black-Scholes option pricing model, other thing being equal, different exercise prices result in different option values. Instead of using a single option to derive the ISD formula, this chapter also explores multiple options with different exercises to derive the ISD formulas. Previous studies focus on a single option's expansion. The first term of the expansion is either at 0.5, if expands, at 0, or if, cumulative normal function, at $\pm\sigma\sqrt{T}$ and expands at-the-money. Unfortunately, $\pm\sigma\sqrt{T}$ is not known and therefore the first term of cumulative normal function at $\pm\sigma\sqrt{T}$ needs to expand further and inevitably results in some amount of truncation error due to the remaining terms in the Taylor expansion. In fact, as shown in Sect. 65.4, the error due to the first term in the expansion can be corrected if different exercises prices are used in the derivation. Given different exercises and the corresponding call option values, applying the Taylor expansion to the Black-Scholes pricing model on different exercise price produces a more accurate first term in the expansion than that in previous studies. Based on two call option prices for different exercise prices and on the Black-Scholes's model, we apply Taylor's formula to (1) both call option values, (2) the inverse cumulative normal functions. Ignoring the remainder terms of the Taylor's formula results in a quadratic equation of the ISD. Solving this equation produces the second formula to estimate the ISD. This formula can estimate the ISD for a wide range of the deviation between the underlying asset price and the exercise prices. Extending the same procedure to a third call option value, we derive the third formula to estimate the ISD.

The chapter is organized as follows. Section 65.2 presents an exact closed-form solution for the ISD if the underlying asset price equals the present value of the exercise price. Section 65.3 presents a formula for the ISD based on a single call option. Section 65.4 derives two formulas to estimate the ISD using two and three call options on the same underlying asset. Section 65.5 presents a simulation to compare the accuracy of the estimated ISD from different formulas. Section 65.6 summarizes our findings.

65.2 An Exact Closed-Form Solution for the Implied Standard Deviation: A Special Case

The call option pricing model derived by Black and Scholes (1973) and Merton (1973) is given by

⁵ Although including the third (or higher) order in the Taylor expansion could improve the accuracy further but it would eliminate the applicability of the quadratic formula of ISD.

$$C = S \left[N \left(\ln(S/K) / (\sigma\sqrt{T}) + \sigma\sqrt{T}/2 \right) \right] - K \left[N \left(\ln(S/K) / (\sigma\sqrt{T}) - \sigma\sqrt{T}/2 \right) \right], \quad (1)$$

where, C is the call premium, S is the underlying asset value, the present value of the exercise price K is equal to Ee^{-rT} . E is the exercise price, r is the instantaneous risk-free rate, T is the time to the maturity, σ is the standard deviation of the underlying asset rate of return on annual basis, and $N(x)$ is the standard cumulative normal distribution function up to x .

Under the condition that the stock price equals the present value of the exercise price (i.e., $S=K=Ee^{-rT}$), Brenner and Subrahmanyam (1988) derive the following simple formula to compute the implied standard deviation.

$$\sigma\sqrt{T} = \sqrt{2\pi}C/S. \quad (2)$$

This formula is derived by applying Taylor series expansion at zero in the B/S model up to the first-order term, and the higher order terms are ignored. Since the higher order terms are omitted, Eq. (2) is an approximation rather than an exact solution. The accuracy of the ISD calculated by Eq. (2) depends not only on the deviation of the stock price S from the present value of exercise price K but also on the magnitude of the standard deviation $\tilde{\sigma}$.

However, when the underlying asset price equals the present value of the exercise price, the B/S model becomes⁶

$$\begin{aligned} C &= S [N(\sigma\sqrt{T}/2) - N(-\sigma\sqrt{T}/2)] \\ &= S [1 - 2N(-\sigma\sqrt{T}/2)] \\ &= S [2N(\sigma\sqrt{T}/2) - 1]. \end{aligned} \quad (3)$$

Thus, an exact closed-form solution for the ISD⁷ is

$$\begin{aligned} \sigma\sqrt{T} &= -2N^{-1}[(S - C)/(2S)] \\ &= 2N^{-1}[(S + C)/(2S)]. \end{aligned} \quad (4)$$

Since the cumulative normal distribution is a monotonic increasing function, its inverse cumulative normal functions in Eq. (4) must exist and be unique. Therefore, Eq. (4) must be an exact and unique closed-form solution for the ISD if the present value of the exercise price equals the underlying asset price. It is noteworthy that the Taylor expansion plays no role in deriving Eq. (4) and there is no truncation of the remainder terms. Hence, when the assumption $S=K=Ee^{-rT}$ holds, using Eq. (4) to

⁶Note that $N(-\sigma\sqrt{T}/2) = 1 - N(\sigma\sqrt{T}/2)$.

⁷Given $S=K$, the BS model becomes $C = S[N(\sigma\sqrt{T}/2) - N(-\sigma\sqrt{T}/2)] = S[N(\sigma\sqrt{T}/2) - (1 - N(\sigma\sqrt{T}/2))] = -S + 2SN(\sigma\sqrt{T}/2)$, solve the ISD yields Eq. (4).

calculate the ISD results in the true ISD and as there is no estimation error. Previous formulas are approximate because they are derived by truncating the remainder terms in the Taylor formula. The estimate error always exists. Equation (4) is the best formula for ISD and dominates all other formulas if the $S = K$. Equation (4) is useful to calculate the exact ISD in this special case because the popularity of the Excel, in which the inverse standard cumulative normal function is imbedded.⁸

Given Eq. (3), the sensitivity of the call option of the exercise price and the hedge ratio at the $S = K$ are determined by

$$\partial C / \partial K = -N\left(-\sigma\sqrt{T}/2\right) = -1/2 + C/(2S), \quad (5)$$

$$\partial C / \partial S = -N\left(\sigma\sqrt{T}/2\right) = 1/2 + C/(2S). \quad (6)$$

Equation (5) is the sensitivity of the call option value from the striking price at $K = Ee^{-rT} = S$. Eq. (6) is the hedge ratio (delta) at the underlying asset price S equals the present value K of exercise price. Equation (5) shows that the $\partial C / \partial K$ must be greater than $-1/2$, while Eq. (6) demonstrates that the delta must be greater than $1/2$. Equations (5) and (6) provide a closed-form and exact solutions for the $\partial C / \partial K$ and delta. Given the closed-form and exact solution of the ISD in Eq. (4), it is possible to find all of the closed-form and exact solutions for all of the Greeks of options at $S = K$. The following theorem summarizes the results of Eqs. (4), (5), and (6).

Theorem *Under the Black and Scholes call option pricing model, if underlying asset price S equals to the present value K of exercise price, then there is an exact closed-form solution for the ISD and is given by Eq. (4). Further, $\partial C / \partial K$ and the hedge ratio at $S = K$ are given by Eqs. (5) and (6) respectively.*

65.3 An Implied Standard Deviation Formula Under a Single Call Option

This section derives a formula to estimate the ISD by applying the Taylor series expansion on a single call option. We show that the improved formula for ISD derived by Corrado and Miller (1996) can be improved further without any replacements. Applying Taylor formula to both cumulative normal distributions in Eq. (1) at $L = \ln(S/K)/(\sigma\sqrt{T})$ yields

⁸The function used to calculate the inverse cumulative normal function is named as “NORMSINV” in f_x of the Excel.

$$\begin{aligned}
N(L + \sigma\sqrt{T}/2) &= N(L) + N'(L)\sigma\sqrt{T}/2 + N''(L) \\
&\quad \left(\sigma\sqrt{T}/2\right)^2/2 + e_1 = N(L) + N'(L)\left(\sigma\sqrt{T}/2\right) \\
&\quad [1 - \ln(S/K)/4] + e_1
\end{aligned} \tag{7}$$

and

$$\begin{aligned}
N(L - \sigma\sqrt{T}/2) &= N(L) + N'(L)\sigma\sqrt{T}/2 + N''(L) \\
&\quad \left(\sigma\sqrt{T}/2\right)^2/2 + e_2 = N(L) + N'(L)\left(\sigma\sqrt{T}/2\right) \\
&\quad (1 + \ln(S/K)/4) + e_2
\end{aligned} \tag{8}$$

where, e_1 and e_2 are the remainder terms of the Taylor's formulas. Equations (7) and (8) are obtained by the fact that $N''(x) = -N'(x)x$.

Given $N(0) = 1/2$, $N'(0) = 1/C\sqrt{2\pi}/S$, $N'''(0) = -N'(0)$, $N''(0) = 0 = N''''(0)$. We apply the Taylor's formula to the $N[\ln(S/K)/(\sigma\sqrt{T})]$ and $N'[\ln(S/K)/(\sigma\sqrt{T})]$ in Eqs. (7) and (8) at 0

$$\begin{aligned}
N(L) &= N(0) + N'(0)(L) + N''(0)(L)^2/2 + e_3 \\
&= 1/2 + L/\sqrt{2\pi} + e_3
\end{aligned} \tag{9}$$

$$\begin{aligned}
N'(L) &= N'(0) + N''(0)(L) + N'''(0)(L)^2/2 + e_4 \\
&= 1/\sqrt{2\pi} - L^2/(2\sqrt{2\pi}) + e_4
\end{aligned} \tag{10}$$

where e_3 and e_4 are the remainders of the Taylor's formulas. Substituting Eqs. (7), (8), (9), and (10) into Eq. (1), dropping all of remainder terms, Eq. (1) becomes

$$\begin{aligned}
C &= (S/K)/2 \\
&\quad + L\left[(S-K)\left(1 + [\ln(S/K)/4]^2\right) - \ln(S/K)(S+K)/4\right]/\sqrt{2\pi} \\
&\quad + \left(\sigma\sqrt{T}/2\right)[S+K - \ln(S/K)(S-K)/4]/\sqrt{2\pi}.
\end{aligned} \tag{11}$$

Equation (11) is a quadratic equation of $\sigma\sqrt{T}$ and can be rewritten as

$$\begin{aligned}
\sigma^2 T[4(S+K) - (S-K)\ln(S/K)] - 4\sigma\sqrt{T}\sqrt{2\pi}(2C - S + K) \\
+ 8\ln(S/K)\left[(S-K)\left(1 + (\ln(S/K)/4)^2\right) - (S+K)\ln(S/K)/4\right] = 0
\end{aligned} \tag{12}$$

Solving $\sigma\sqrt{T}$ from Eq. (12) yields

$$\sigma\sqrt{T} = \left[b \pm \sqrt{b^2 - ac}\right]/a \tag{13}$$

Where $a = 4(S + K) - (S - K)\ln(S/K)$, $b = 2\sqrt{2\pi}(2C - S + K)$, $c = 8\ln(S/K) [(S - K)(1 + (\ln(S/K)/4)^2) - (S + K)\ln(S/K)/4]$. A merit of Eq. (13) is to circumvent the ad hoc substitution present in CM and improve the accuracy of the ISD's estimation.

Since Eqs. (9) and (10) are the Taylor series at the zero point and the remainder terms e_1 to e_4 in Eqs. (7), (8), (9), and (10) are omitted in deriving Eq. (11), the ISD calculated by Eq. (13) is not an exact formula. Therefore, the effectiveness of using Eq. (13) to estimate the ISD depends on the deviation of the underlying asset price from the present value of exercise price. The merits of this formula are: (Eq. 1) it is derived without arbitrary substitution (see Sect. 65.7) as Corrado and Miller (1996), and (Eq. 2) it provides a more accurate estimate of ISD than that of Corrado and Miller. As shown in the figures in Sect. 65.5, the estimated ISD based upon Eq. (13) is more accurate and dominates the formula provided by Corrado and Miller (1996). Section 65.5 addresses in detail the accuracy of the ISD calculated by Eq. (13) and that of the improved formula derived by CM.

65.4 Formulas for Implied Standard Deviation Under Different Exercise Prices

In this section we derive two formulas to estimate the ISD from call options with different exercise price. The derived formulas can apply to a wide range of exercise prices that deviate from the underlying asset price. The advantage from using options with different exercise prices in this section is to obtain a no error's first term in the expansion, thus, gives more accurate formulas for ISD. The first term in Taylor expansion usually provides the major component of the true function, the remaining terms provide correction to adjust and approximate the true function. For example, if the call option value C_1 with exercise price E_1 is expanded at the exercise price E_2 , the first term of the Taylor series expansion is the given option value C_2 on E_2 . Similarly, the first term of the Taylor expansion on option value C_2 on E_2 at exercise price E_1 is C_1 . Since both C_1 and C_2 are given, there is no need to expand the first terms again.

Consider two call options C_1 and C_2 on the same time to the maturity with prices of E_1 and E_2 (where $E_1 < E_2$), their present value are K_1 , and K_2 , respectively. Applying Taylor's expansion to Eq. (1) at K_2 for C_1 and at K_1 for C_2 respectively yield,

$$C_1 = C_2 - N\left(\ln(S/K_2)/(\sigma\sqrt{T}) - \sigma\sqrt{T}/2\right)(K_1 - K_2) + \varepsilon_1, \quad (14)$$

$$C_2 = C_1 - N\left(\ln(S/K_1)/(\sigma\sqrt{T}) - \sigma\sqrt{T}/2\right)(K_2 - K_1) + \varepsilon_2. \quad (15)$$

Here ε_1 and ε_2 are the remainder terms of C_1 at K_2 and C_2 at K_1 from Eq. (1). Dividing both sides of Eqs.(14) and (15) by $(K_2 - K_1)$ and simple manipulations

produce the same left hand side of $(C_1 - C_2)/(K_2 - K_1)$. Then applying the inverse function of cumulative normal function on both sides and after using the Taylor's formula yields the following equations

$$\begin{aligned} N^{-1}[(C_1 - C_2)/(K_2 - K_1)] \\ = \ln(S/K_1)/(\sigma\sqrt{T}) - \sigma\sqrt{T}/2 + \eta_1, \end{aligned} \quad (16)$$

$$\begin{aligned} N^{-1}[(C_1 - C_2)/(K_2 - K_1)] \\ = \ln(S/K_2)/(\sigma\sqrt{T}) - \sigma\sqrt{T}/2 + \eta_2, \end{aligned} \quad (17)$$

where η_1 and η_2 are the remainder terms of Taylor's formulas derived from Eqs.(14) and (15) respectively. After dropping the remainder terms, ISD solved individually from either Eq. (16) or (17) could be more volatile than that solved simultaneously with both equations. Combining Eqs.(16) and (17) could cause the effects of remainder term ($\eta_1 + \eta_2$) to partially offset and thus, reduce their variability. Dropping the remainder term ($\eta_1 + \eta_2$), a simple manipulation produces the following equation

$$\begin{aligned} (\sigma\sqrt{T})^2 + 2N^{-1}[(C_1 - C_2)/(K_2 - K_1)](\sigma\sqrt{T}) \\ - \ln[S^2/(K_1K_2)] = 0. \end{aligned} \quad (18)$$

Equation (18) is a quadratic equation of the ISD and thus can be solved as,

$$\begin{aligned} \sigma\sqrt{T} = -N^{-1}[(C_1 - C_2)/(K_2 - K_1)] \\ \pm \sqrt{[N^{-1}((C_1 - C_2)/(K_2 - K_1))]^2 + \ln(S^2/K_1K_2)}. \end{aligned} \quad (19)$$

There are two solutions for the ISD in Eq. (18), yet the ISD must only take one of them. It is clear that, if $(C_1 - C_2)/(K_2 - K_1) > 1/2$, the inverse of cumulative normal function on it must be non-negative. Consequently, ISD must take the value with the positive sign in Eq. (19), or else ISD will become a negative, which violates the positive requirement for ISD. On the other hand, if $(C_1 - C_2)/(K_2 - K_1) < 1/2$ and $K_1 < S < K_2$, the ISD must take the positive sign in Eq. (19) too. The negative sign only apply when S is less than K_1 . It is clear that if stock price is less than the lower exercise K_1 (i.e., then both call options are out-of-the-money) and if we had chosen the value with the plus sign in Eq. (19), ISD calculated by Eq. (19) will be overstated.

Furthermore, if K_2 approaches K_1 , then $(C_1 - C_2)/(K_2 - K_1)$ in Eq. (19) tends to $-\partial C/\partial K$ and $\ln(S^2/K_1K_2)$ tends to zero. Given positive ISD, ISD calculated by Eq. (19) must converge to $-2N^{-1}(-\partial C/\partial K)$ (see Eq. 5). Convergence implies that the ISD in Eq. (19) must take the negative sign for the exercise prices. Note that the ISD in Eq. (19) is not an exact solution because the remainder terms η_1 and η_2 were dropped out in the derivation of Eq. (18).

Equation (19) provides a formula to calculate the ISD. All of the parameters on the right hand sides of Eq. (19) are given. A merit of this formula is that a sufficient condition to calculate ISD by Eq. (19) only requires that there existed any two consecutive call option values with different exercise prices. Although it may not always hold that the call option value at $S=K$, it is easier to obtain any consecutive two traded call option values and their exercise prices in practice. In particular, Eq. (19) is applicable to a wide range of exercise prices if their call options are observable. Of course, its accuracy will depend on the magnitude of the deviation between these two exercise prices. The closer are these two exercise prices, the lower is the remainder of the Taylor Formulas in Eqs.(14) and (15). Consequently, the ISD estimated by Eq. (19) is more accurate than in previous models. In particular, both formulas by Corrado and Miller (1996) and by Li (2005) are only valid when calculated at near at-the-money. Outside this range, both formulas' performance is poor if not applicable.

Similar to Eq. (18), if there is a third call option C_3 with the discounted exercise price K_3 , then the following Eq. (20) must hold for K_2 , K_3 and C_2 , C_3 .

$$\begin{aligned} & \left(\sigma\sqrt{T}\right)^2 + 2N^{-1}[(C_2 - C_3)/(K_3 - K_2)]\left(\sigma\sqrt{T}\right) \\ & - \ln [S^2/(K_2K_3)] \\ & = 0. \end{aligned} \quad (20)$$

Given the constant variance assumption in Black and Scholes option model, the following Eq. (21) is thus derived by subtracting Eq.(20) from Eq. (18).

$$\sigma\sqrt{T} = \ln(K_3/K_1)/[2(N^{-1}((C_1 - C_2)/(K_2 - K_1)) - N^{-1}((C_2 - C_3)/(K_3 - K_2)))]. \quad (21)$$

An advantage of using Eq. (21) rather than Eq. (19) to estimate the ISD is to circumvent the sign issue that appears in Eq. (19). However, a drawback of using Eq. (21) is that there must exist at least three instead of two call options for Eq. (19). Equation (21) provide a simple formula to calculate ISD because all options values and exercises price are given and the inverse function of the standard cumulative normal function also available in the Excel spreadsheet. Instead of calculating a complex function of cosine and inverse cosine function in the formula in Li (2005), Eq. (19) provides a more convenient approach to calculate ISD./Furthermore, the Li's formula suffers from limited applicable range as stated before. The accuracy of the ISD calculated by Eqs.(19) and (21) will be addressed in Sect. 65.5.

65.5 Accuracy of the ISD Models: Simulation Results

This section uses simulation to compare the accuracy of the estimated ISDs by different formulas. To conduct the simulations, we assume the following parameter values: the underlying asset price (S) is \$100, continuous annual risk-free rate (r) is 6%, time to maturity (T) is at: 1 month, 3 months, 1 year and 2 years, and the

constant standard deviation σ is to vary from: 10%, 20%, 50%, 75%, and 80%. Under these assumptions, we use the Black and Scholes' option model, Eq. (1) to calculate the theoretical call option values for different exercise prices. Finally, given the theoretical call option value and the assumed risk-free rate, underlying asset price, time to the maturity, and the exercise prices, we calculate the ISD from different formulas. We then compare the true ISD and the calculated ISD to gauge the accuracy of the formulas for the ISD.

We start with comparing our three formulas (using one, two, and three options) with Corrado and Miller's two formulas (original and improved) for options with 3 months maturity. In Fig. 1, we estimate the implied standard deviations using the parameters above. From these implied standard deviations, we calculate percentage estimation errors. Figure 1a, b plot the estimation errors of the two Corrado and Miller (CM) formulas under different asset prices with respect to exercise price of \$100, i.e., moneyness of the options. We find the original CM has limited range of applicability: (a) It does not produce a positive ISD for options with small true standard deviations, or (b) for options that are far out or in the money. More serious is the rather large estimation errors that could be as large as 30–50% in underestimation. The improved CM manages to reduce underestimation error by several orders of magnitude, but nevertheless still suffer from limited range of applicability.

On the other hand our formula, using a single option (Eq. 13, Fig. 1c), manage to produce accurate estimates for some options with small standard deviations.

By adding one more option, (Eq. 19, Fig. 1d), not only is applicable to all options, the estimated ISDs reduce absolute estimation errors by several order of magnitude.

However, Eq. (19) produces both positive and negative estimation error (over and under estimate), and there are kinks in the error functions for options in the neighborhood of the exercise price. Fortunately, this problem is solvable as demonstrated in the error functions for Eq. (21), Fig. 1e, with three options. We find these functions are well behaved for a wide range of moneyness of the options. Although ISDs calculated from Eq. (21) tend to overestimate true standard deviation, the estimation errors are extremely small, only fraction of a percent. Furthermore, it has the widest range of applicability to options with different standard deviations. The accuracy of this formula, although ready high, improves as the price of the underlying asset increases, i.e., high accuracy for out of money options but extremely accurate for deep in the money options.

We also compare the estimation errors of these formulas under shorter maturity, 1 month (Fig. 2a–c), and longer maturity of 1 year (Fig. 3a–e). We find that discussions above generally hold: (a) CM still suffers from very large estimation errors and rather limited range of applicability with respect to both moneyness and true standard deviations. (b) Equation(13) performs as well if not better than CM improved, but has the benefit of not having to make arbitrary substitutions. (c) Equation(19) makes a marked improvement especially for out of money options, however, it still has undesirable kinks for in the money options. (d) Equation(21) consistently delivers the best overall results, it produces estimation errors around 1% or less for out of money options, and practically negligible for out of money options for all values of standard deviations used in the simulations.

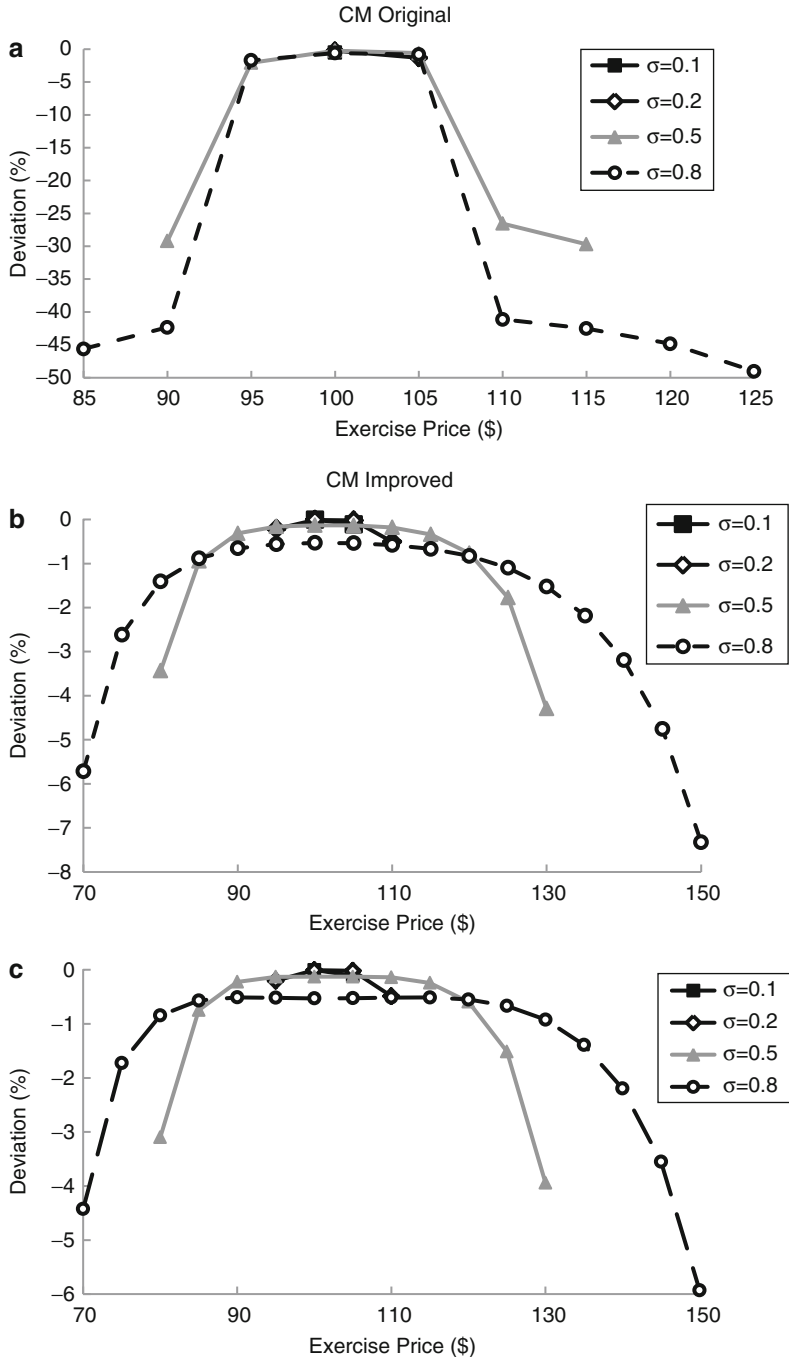


Fig. 1 (continued)

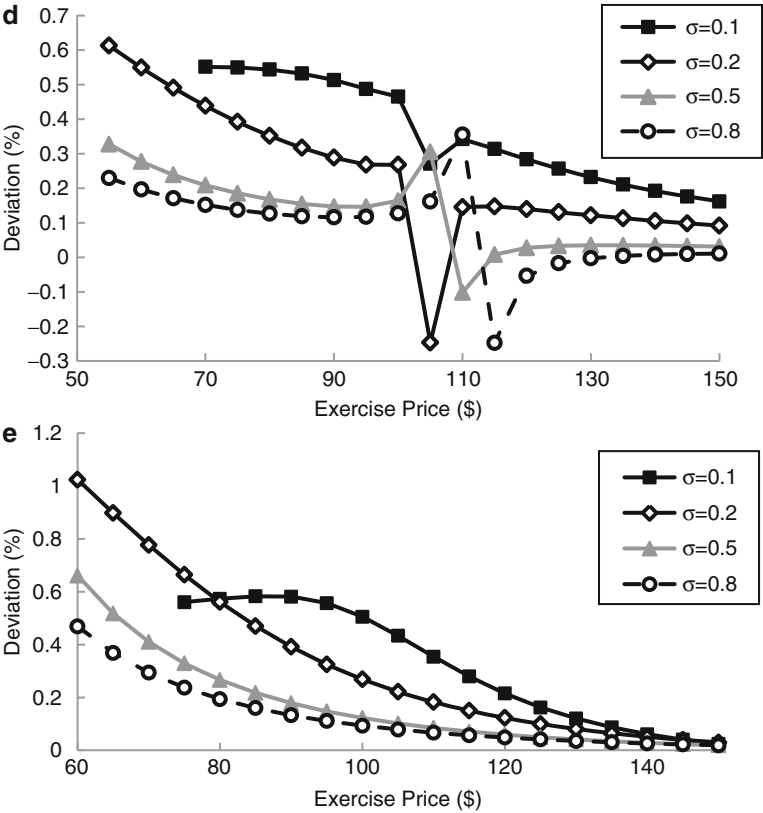


Fig. 1 A comparison of various formula to compute implied standard deviation: simulation results for 3 month options under different true standard deviation and moneyness (distance between current and exercise price). The parameters are: the underlying asset price (S) is \$100, continuous annual risk-free rate (r) is 6%, time to maturity (T) is at: 1 month, 3 months, 1 year and 2 years, and the constant standard deviation σ is to vary from: 10%, 20%, 50%, and 80%. (a) ISD is calculated from Corrado Miller (CM, 1996): The Y axis gives the prediction error as estimated ISD minus true ISD divided by true ISD, and X axis gives the distance of exercise price from market price at \$100. (b) ISD calculated from the improved version of Corrado and Miller (2004). (c) ISD calculated from Eq. (13). (d) ISD calculated from Eq. (19). (e) ISD calculated from Eq. (21)

Finally, we examine the effect of maturity on estimation error. Here, we concentrate on Eq. (2) as it is the best performing formula. We find that estimated ISDs under the formula (Fig. 4a–c) improve with longer maturity, greater true standard deviation, and greater moneyness, i.e., more in the money options. For options with maturity over a year, estimated errors under Eq. (21) at around a tenth of 1%, is for all purpose not of practical significance.

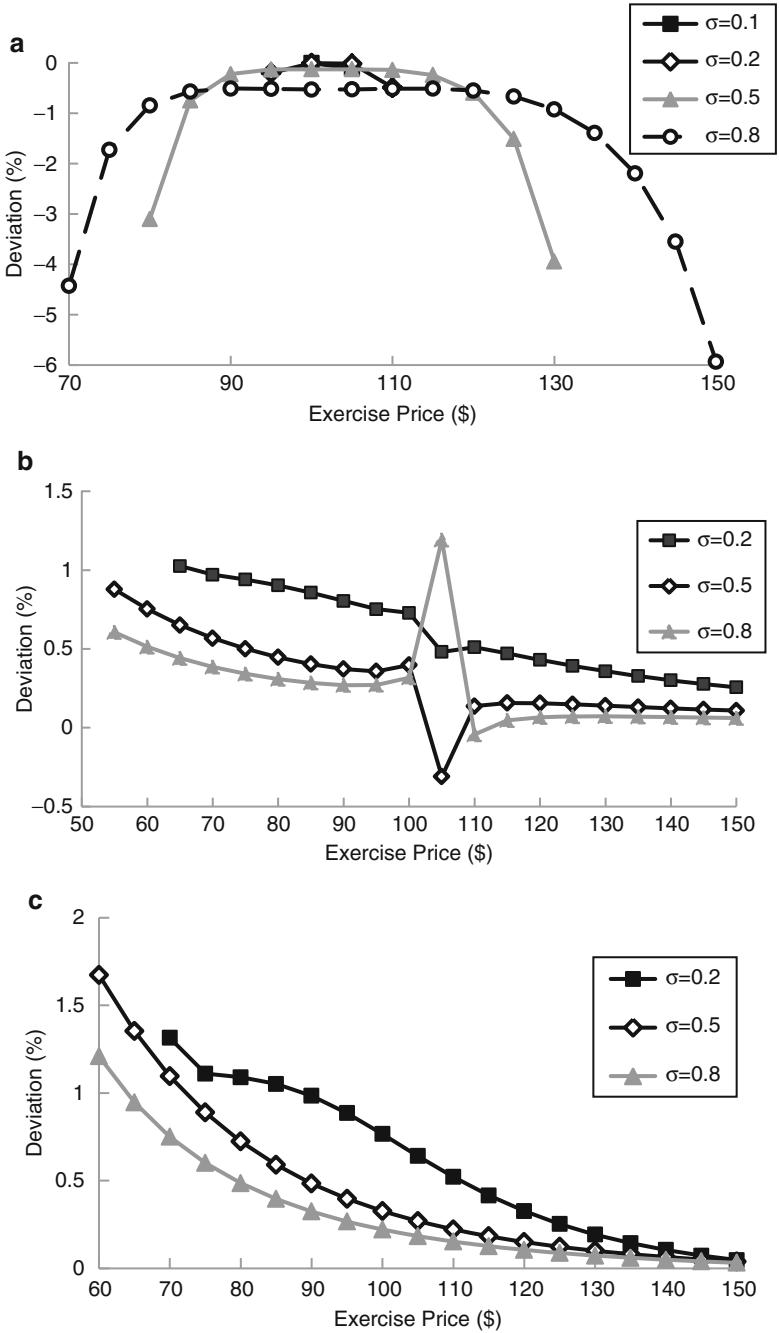


Fig. 2 A comparison of various formula to compute implied standard deviation: simulation results for 1 month options under different true standard deviation and moneyness (distance between current stock price and different exercise prices). Parameters same as in Fig. 1. (a) ISD calculated from Eq. (13). (b) ISD calculated from Eq. (19). (c) ISD calculated from Eq. (21)

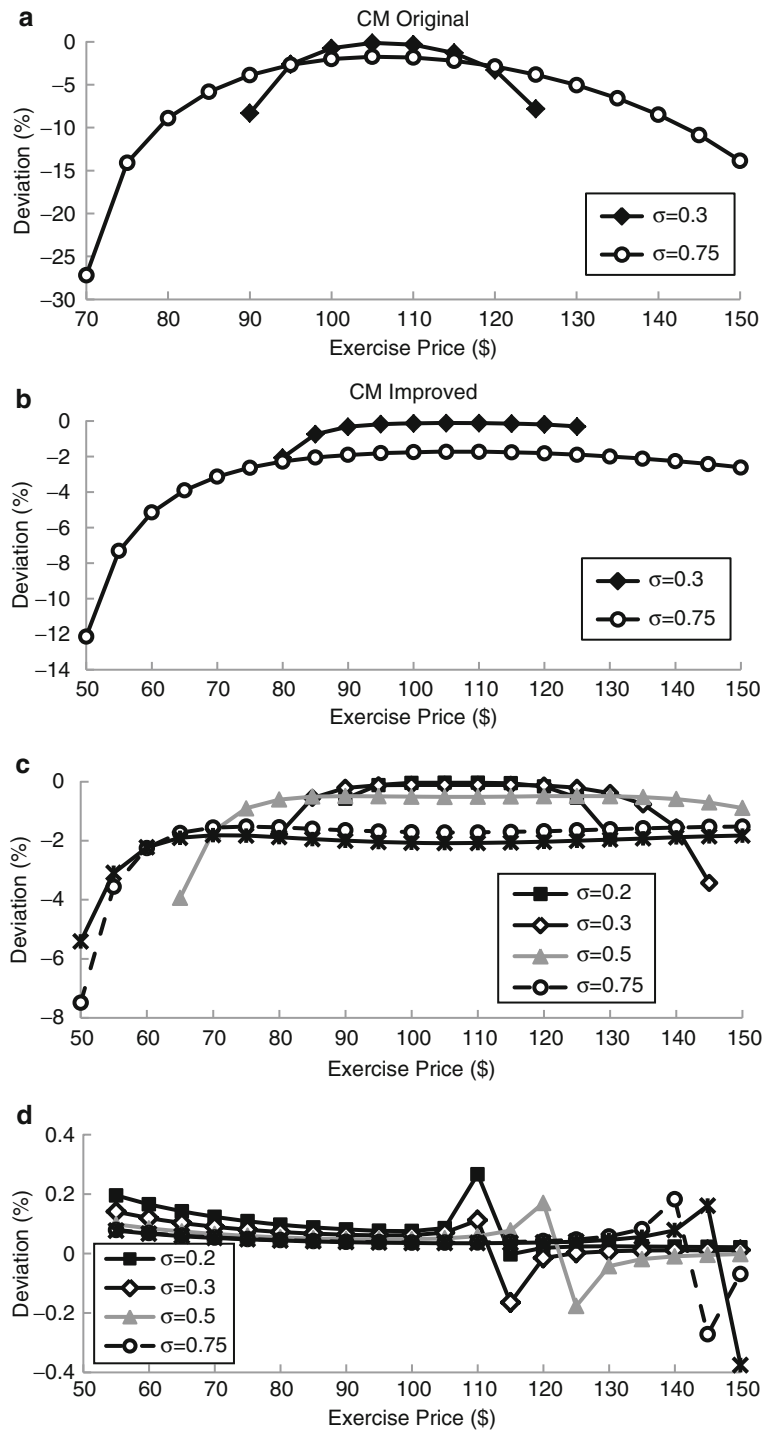


Fig. 3 (continued)

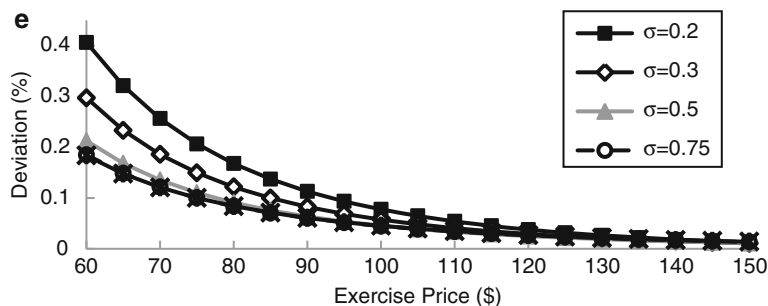


Fig. 3 Simulation results for 1 year maturity options. Parameters are the same as in Fig. 1. (a) ISD calculated from Corrado Miller's original. (b) ISD calculated from Corrado and Miller Improved. (c) ISD calculated from Eq. (13). (d) ISD calculated from Eq. (19). (e) ISD calculated from Eq. (21)

65.6 Conclusion

This chapter derives an exact closed-form solution for the implied standard deviation under the condition that the underlying asset price equals the present value of the exercise price. Given this closed-form solution for the implied standard deviation, there are closed-form solutions for the delta (hedge ratio) and the other Greeks of the option.

When we allow the underlying asset price to deviate from the present value of the exercise price, we derive three formulas to estimate the implied standard deviation. Applying the Taylor series expansion to the Black-Scholes option pricing model up to the second order, we derive the first formula to estimate the implied standard deviation. The ISD estimated by this formula is accurate for the near at-the-money call options. We then use two call options at different exercise prices to derive the second formula. This formula further improves the accuracy of the estimated ISD. We further extend this approach to include a third option to derive an even more accurate third formula. We show by simulation that this formula produce consistently accurate ISD for all parameter values in maturity, moneyness, and true standard deviations.

65.7 Appendix

Based upon Black and Scholes' option pricing model, Brenner and Subrahmanyam (1988) derive the following formulae of ISD under the condition that the stock price S equal to the present value K of the exercise price E :⁹

⁹This Appendix is suggested by an anonymous referee.

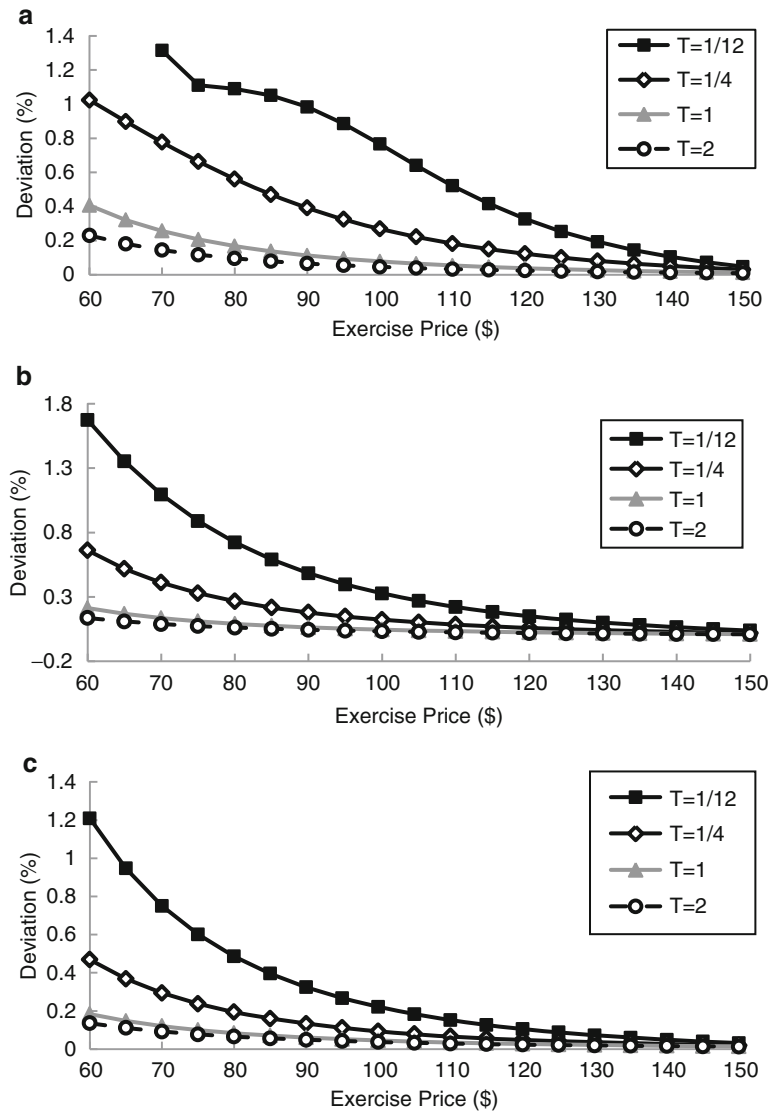


Fig. 4 The effect of maturity on estimation results: ISD estimated from Eq. (21)

$$\sigma\sqrt{T} = \sqrt{2\pi}C/S \tag{22}$$

Corrado and Miller (1996) expand the cumulative normal distribution in Black and Sholes model up to the third order and omit the rest of reminders. The resulting Taylor expansion of option pricing model on the zero point results a quadratic equation of the ISD. They then solve for the ISD from the quadratic equation and

derive the following formulae (23) of ISD (their Eq. (6) in Corrado and Miller 1996, detail see page 597).

$$\sigma\sqrt{T} = \sqrt{2\pi}[(C - (S - K)/(S + K)] \\ + \sqrt{2\pi[(C - (S - K)/2)/(S + K)]^2 - 2(S - K) \ln(S/K)/(S + K)}. \quad (23)$$

Corrado and Miller stated “The accuracy of the Eq. (6) can be significantly improved by minimizing its concavity as follows. First, for simplicity only, we substitute the logarithmic approximation $\ln(S/X) \approx 2(S - X)/(S + X)$ into Eq. (6) and then replaced the value “4” with parameter α to obtain this restatement of the quadratic formula.”

$$\sigma\sqrt{T} = \sqrt{2\pi}[(C - (S - K)/2)/(S + K)] \\ + \sqrt{2\pi[(C - (S - K)/2)/(S + K)]^2 - \alpha[(S - K)/(S + K)]^2}. \quad (24)$$

This α is called tweaking parameter by Corrado and Miller (2004). To minimize the concavity of their Eq. (7) (here the Eq. (24) at the stock price equal to the present value X of exercise price, they derive a formula for α in terms normal density function and cumulate normal function with ISD. They solve the α (α is a function of unknown variable of ISD, detail see their Eqs.(8) and (9) in page 598) and then substitute $\alpha = 2$ to obtain the following formula of ISD Eq. (25)(the Eq. 10 in Corrado and Miller):

$$\sigma\sqrt{T} = \sqrt{2\pi} \left[C - (S - K)/2 + \sqrt{(C - (S - K)/2)^2 - (S - K)^2/\pi} / (S + K) \right]. \quad (25)$$

As shown by Corrado and Miller, their Eq. (10) (here Eq. 25) is more accurate than their Eq. (6) (here Eq. 23).

References

- Ang, J.S., G.D. Jou, and T.Y. Lai. June 2009. Alternative formulas to compute implied standard deviation. *Review of Pacific Basin Financial Markets and Policies* 12: 159–176.
- Beckers, S. 1981. Standard deviation implied in option prices as predictors of future stock price variability. *Journal of Banking and Finance* 5: 363–381.
- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–654.
- Brenner, M., and M.G. Subrahmanyam. 1988. A simple formula to compute the implied standard deviation. *Financial Analysts Journal* 44: 80–83.
- Chance, D. 1996. A generalized simple formula to compute the implied volatility. *Financial Review* 31: 859–867.
- Corrado, C.J., and T.W. Miller. 1996. A note on a simple, accurate formula to compute implied standard deviations. *Journal of Banking and Finance* 20: 595–603.

- . 2004. *Tweaking implied volatility*, Working paper. Auckland: Massey University–Albany.
- Garman, M.S., and M.J. Klass. 1980. On the estimation of security price volatility from historical data. *Journal of Business* 53: 67–78.
- Hallerback, W.G.P.M. 2004. *An improved estimator for Black-Scholes-Merton implied volatility*, ERIM report series. Rotterdam: Erasmus University.
- Lai, T.Y., C.F. Lee, and A.L. Tucker. 1992. An alternative method for obtaining the implied standard deviation. *Journal of Financial Engineering* 1: 369–375.
- Latane, H.A., and R.J. Rendleman. 1976. Standard deviation of stock price ratios implied by option premia. *Journal of Finance* 31: 369–382.
- Li, S. 2005. A new formula for computing implied volatility. *Applied Mathematics and Computation* 170: 611–625.
- Manaster, S., and G. Koehler. 1982. The calculation of implied variances from the Black–Scholes model: A note. *Journal of Finance* 37: 227–230.
- Merton, R.C. 1973. The theory of rational option pricing. *Bell Journal of Economics and Management Science* 4: 141–183.
- Schwert, G.W. 1989. Why does stock market volatility change over times? *Journal of Finance* 44: 1115–1154.



Securities Transaction Taxes: Literature and Key Issues

66

Anna Pomeranets

Abstract

The main scope of this chapter is to review the literature and key issues associated with securities transaction taxes (STTs). Despite the use of STTs around the globe, the theoretical and empirical literature on the impact an STT has on liquidity and volatility is mixed.

If an STT is not appropriately designed, it could interfere with the smooth functioning of financial markets, lead to informational inefficiency, arbitrage, tax evasion and double taxation. Effective implementation of STTs therefore requires cross-jurisdictional coordination, controls on cross-border transactions and carefully constructed enforcement.

Keywords

Tobin tax · Transaction costs · Transaction tax · Transfer tax · Volatility

66.1 Introduction

Financial markets are able to convert the underlying demands of investors into financial transactions. The imposition of a security transaction tax (STT), a tax on the purchase or sale of a financial security, affects this conversion. Proponents of STTs argue that such a tax, in addition to raising revenues for public services, may reduce excessive trading and market volatility because it creates a disincentive for short-term speculative trading by placing a financial burden on these transactions. On the other hand, because financial markets efficiently allocate resources in the economy, STTs reduce the ability of markets to do so and therefore reduce welfare.

A. Pomeranets (✉)
Bank of Canada, Ottawa, ON, Canada
e-mail: APomeranets@bank-banque-canada.ca

As past events have shown, STTs are not necessarily without risks and can have an unintended impact on financial markets.¹

Despite the use of STTs around the globe, many aspects of their impact on financial markets are frequently debated by academics, politicians and policymakers.² In addition, the literature on the impact that an STT has on volatility and liquidity is inconclusive. This chapter reviews the existing literature and key issues associated with STT.

66.2 Theoretical Literature

Views on the impact STTs have on financial markets are divided. The earliest proponents of STTs, John Maynard Keynes and James Tobin, voiced the following rationale for the benefits of transaction taxes. First, the resources that financial markets require outweigh the role they play in economic welfare. The amount of capital that changes hands in financial markets exceeds the value of the fundamental transaction. Keynes (1936) contends that chasing short-term returns while potentially profitable, is a zero-sum game in terms of economic welfare. Because one investor's gain is another one's loss and trading utilizes resources, the value added through trading is negative. As a result, imposing an STT may increase welfare by reducing wasted resources. Second, because trading is speculative by nature, it potentially contributes to financial instability when trades are driven by short-term capital gains and not fundamental information. An STT could curtail short-term speculation, thereby reducing wasted resources, market volatility and asset mispricing. Finally, James Tobin in 1978 proposed a foreign exchange transaction tax to be levied globally in order to minimize cross-border capital flows that impair governments' efforts to control total demand. Thus, a tax on foreign exchange transactions would be a fiscal form of control that would make some cross-border dealings unprofitable, particularly in the short-term.

These original arguments were the grounds for many proposals and implementations of STTs around the world, and have ignited a long-lasting debate among academics, politicians and policymakers. This section reviews the literature around the impact an STTs has on volatility, liquidity and volume.

66.2.1 Volatility

Many advocates have furthered the debate in recent years, leading to an array of results. To date, there are three groups of papers that propose three contradicting theories on the impact an STT has on volatility.

¹See Umlauf (1993) and Cortez and Vogel (2011).

²STTs have been used in the U.S., U.K., Taiwan, Hong Kong and Sweden, among others. For more details see Matheson (2011).

The first group of papers, consistent with Keynes and Tobin, suggest that an STT will decrease volatility because it targets short-term noise traders. Stiglitz (1989) and Summers and Summers (1989) maintain that an STT would curb short-term speculation, thereby reducing exhausted resources, security mispricing and market volatility. Stiglitz (1989) theorizes that the tax targets short-term noise traders whose trades make markets more volatile. By using an STT to discourage noise traders from trading, volatility should be reduced. Similarly, Summers and Summers (1989) evaluate the costs and benefits of an STT. They suggest that excessive speculation increases volatility and conclude that the benefits of an STT outweigh its costs. Summers and Summers propose that the tax will reduce the activity of noise traders, resulting in the transaction price deviating less from the efficient price and thus causing a decrease in *excess* volatility and an increase in market quality.

Opponents of STTs focus on its potential to distort financial markets by reducing volume and liquidity, increasing price volatility and causing inefficient price discovery (see Amihud and Mendelson 1992). They contend that an STT can lead to lower asset prices, increased cost of capital for businesses, lower returns to savings and widespread tax evasion. The second group of papers is consistent with the views of opponents and proposes an ambiguous impact of the tax on volatility because the tax will affect both informed traders and noise traders. For example, Schwert and Seguin (1993) suggest that there is no evidence of *excess* volatility and since the tax is a burden on all traders, the reduction in trading will not limit the activity of noise traders alone—it will affect liquidity traders and price-stabilizing informed traders as well. They conclude that the STTs' impact on volatility is ambiguous and emphasize the unclear effect it would have on market quality.

The final portion of the theoretical literature concludes that an STT will have a direct effect on volatility because it reduces informed trading. Kupiec (1996) develops a model where an increase in the level of the STT increases excess volatility. Similarly, Amihud and Mendelson (2003) conclude that an STT results in an increase in volatility. They argue that transaction taxes reduce the amount of informed trading thereby widening the gap between the transaction price and the security's fundamental value. For example, when a price deviates from its value, informed traders buy it if the security is undervalued or sell it if it is overvalued; in so doing, the price moves back toward its fundamental value. Since informed traders incorporate transaction costs into their trading decisions, an STT will increase the costs faced by an informed trader and in turn result in larger deviations from fundamental value before profits are large enough to induce informed traders to trade. These larger deviations then result in an increase in volatility.

In summary, the existing theories on the impact an STT has on volatility are inconsistent and can be categorized into three groups. Consistent with advocates of the tax, the first group finds that an STT will result in a decline in volatility because it targets short-term noise traders. The second group contends that an STT will target all investors and therefore its impact on volatility will be ambiguous. The final group, consistent with opponents of an STT, argues that the tax will be accompanied by a rise in volatility. These contradictory theories feed the debate among proponents and opponents of an STT.

66.2.2 Liquidity and Volume

Consistent with the lack of consensus on the impact a securities transaction tax has on volatility, the theoretical literature on the relationship between an STT and liquidity is also inconclusive. Schwert and Seguin (1993) theorize that liquidity would be affected by an STT whereas Stiglitz (1989) theorizes that the impact would be insignificant. Importantly, these papers draw on the indirect effect of trading volume on liquidity to interpret the impact of an STT on liquidity.

Schwert and Seguin (1993) conceive that an STT would indirectly increase transaction costs through the three components of the bid-ask spread: order processing costs, inventory risk and information asymmetry.³ First, since an STT reduces trading volume, the number of trades available for the market-maker to recover his fixed cost declines, thereby increasing the order processing component of the spread. An STT will also increase the order processing component directly due to the cost of collecting the tax. Second, because equity market-makers use derivatives to hedge their risky inventory positions, an STT on derivatives increases the cost of hedging their positions. The increase in the market-maker's cost for hedging increases the inventory risk part of the bid-ask spread. Finally, if an STT reduces the amount of noise trading, as proponents suggest, then the possibility of the market-maker facing an informed trader increases, thereby increasing the information asymmetry element of the bid-ask spread. The authors theorize that an increase in spreads results in an increase in the cost to traders who take advantage of security mispricings, and thus reduces market efficiency. That is, when transaction costs increase, larger and more persistent mispricings are likely.

In contrast, Stiglitz (1989) argues that the impact of an STT would be insignificant by examining the inventory risk component of the bid-ask spread. He contends that although an STT will lead to thinner markets because the tax will discourage both buyers and sellers from trading, the change in liquidity will be insignificant. As markets become thinner, the time period between the chapter of buy and sell orders widens. As a result, market-makers are forced to hold onto risky assets for longer, increasing the inventory risk component of the bid-ask spread. Despite the suggested increase in the bid-ask spread, Stiglitz argues that the impact of an STT on widely traded stocks will be insignificant as the extra minutes or seconds that market-makers hold onto the securities will not yield a significant change.

In summary, the theoretical arguments on the impact STTs have on volatility and liquidity are mixed and inferred from the inverse relationship that an STT has with volume. As a result, there have been numerous efforts to sort out the issues empirically. The next section reviews the findings in the empirical literature.

³The order processing component is part of the fixed cost the market-maker charges for trade execution. The inventory risk component is the market-maker's compensation for holding onto risky assets. The information asymmetry component represents the likelihood that a market-maker is facing an informed trader who has superior knowledge of the asset's fundamental value.

66.3 Empirical Literature

The empirical literature assesses three concerns of STTs: the impact the tax has on volatility, volume and liquidity. This section reviews the findings in the empirical studies around these measures of market quality.

66.3.1 Volatility

The tax is geared toward reducing the amount of speculative trading and thus stock market volatility. Proponents of the tax conjecture that the relationship between an STT and volatility is inverse. However, this proposed relationship is not observed in the empirical literature. Similar to the contradictory results in the theoretical literature on the impact of an STT on volatility, the empirical studies also provide varied results that can be divided into two clusters.

The first group finds a direct relationship between an STT and volatility and contradicts what advocates of the tax propose. Studies such as Umlauf (1993), Jones and Seguin (1997), Green et al. (2000), Hau (2006), and Baltagi et al. (2006) find a direct relationship between STTs and volatility. Umlauf (1993) looks at the Swedish Stock Market in the 1980s and finds that with an increase in an STT, volatility also increases. He finds that during the period with the highest STT rate, daily variances are at their highest levels as well. Jones and Seguin (1997) look at the repeal of fixed commissions on the NYSE and AMEX in 1975 and find that a decrease in commissions is accompanied by a decrease in volatility.⁴ Green et al. (2000) examine the transfer tax in the UK and find that the tax and volatility are directly related. Hau (2006) looks at an increase in transaction costs in the French Stock Market as a result of tick size regulations and finds that as transaction costs increase, volatility also increases. Baltagi et al. (2006) find a significant increase in volatility following an increase in the tax policy from 0.3% to 0.5% in the Chinese Stock Market. These results are consistent with what opponents of the tax argue: an STT has adverse effects on volatility. It should be noted that Jones and Seguin and Hau study changes in transaction costs and not an actual change in the level of an STT. While the abolition of fixed commissions and a tick size change are pseudo-taxes because they are a onetime decrease in the transaction cost, they are not necessarily identical.

The second cluster of papers finds that volatility remains unchanged after a change in the level of an STT. This group of papers is also in stark contrast with what proponents of an STT surmise. For example, Pomeranets and Weaver (2011) examine nine changes in the level of a New York State imposed STT and find no consistent relationship between an STT and volatility. Roll (1989) examines stock return volatility for 23 countries that impose STTs and finds no evidence that volatility is related to transactions taxes. Similarly, Saporta and Kan (1997) find no

⁴The abolition of fixed commissions is a pseudo-tax because it is a one-time decrease in the transaction costs on the New York and American Stock Exchanges.

relationship between the U.K. STT and price volatility, and Hu (1998) finds no reliable relationship between an STT and volatility for changes in the level of an STT in Hong Kong, Japan, Korea, and Taiwan between 1975 and 1994. Bloomfield et al. (2009) create a laboratory market and find that price volatility is higher for periods when noise traders trade actively; however, after an STT is imposed, volatility is not significantly reduced despite the fact that noise traders submit fewer orders. Similarly, Sahu (2008) examines the imposition of an STT on the Indian Stock Market in 2004 and finds that an STT has no impact on volatility.

Both clusters of papers assessing the impact of an STT on volatility are in stark contrast with what advocates of the tax suggest: an STT will result in a decline in volatility. However, the literature does not support this conjecture.

66.3.2 Volume

There is a general consensus in the empirical literature that the relationship between a securities transaction tax and volume is inverse. Jones and Seguin (1997) and Baltagi et al. (2006) find an inverse relationship between volume and an STT. When an STT is imposed, some traders may stop trading. These results are consistent with the theoretical literature on the relationship between volume and a securities transaction tax.

In similar vein, when a transaction tax increases on one exchange, it drives volume to another exchange. Pomeranets and Weaver (2011) find that the New York Stock Exchange market share is inversely related to a New York State imposed STT. Campbell and Froot (1993) look at an STT for 20 different countries and find that an increase in the tax results in a loss of market share domestically and an increase in market share abroad. Umlauf (1993) finds that an increase in a securities transaction tax results in a decline in market share in the domestic country. These results are consistent with what opponents of the tax surmise. An STT drives trading to locations that do not impose an STT.

66.3.3 Liquidity

One main purpose of an organized financial market is to provide liquidity. A financial market should allow traders to buy and sell securities quickly and at low costs. Liquidity can be measured by the bid-ask spread, market depth, immediacy, price impact and resilience. The literature on the impact of stock transaction taxes on liquidity is limited.

Pomeranets and Weaver (2011) measure the impact of an STT on spreads and find an STT has a direct relationship with spread width. That is, imposing or increasing an STT is associated with a wider spread. Pomeranets and Weaver also examine price impact and find that an STT adversely affects market quality.

Bloomfield et al. (2009) study the relationship between liquidity and STTs in a laboratory setting in the context of different types of traders. If noise traders are less

active because of the STT, the inventory risk component of the spread may decrease; the adverse selection cost may increase; or it may remain unchanged. If noise traders trade less, then prices will be less volatile and the inventory risk component of the bid-ask spread will decline, thereby increasing liquidity. However, if noise traders trade less, increasing the probability of informed trading, then the adverse selection costs could increase, resulting in a decline in liquidity. Finally, given a decline in noise trading from an STT, informed traders may trade less aggressively on their information, keeping adverse selection costs unchanged. Bloomfield et al. (2009) use the total price impact as a proxy for liquidity and find no significant impact on the bid-ask spread or price impact.

Empirical examinations face three chief obstacles. First, the effects of taxes on prices and volume are difficult to unravel from market structure and other policy changes that coincide with an STT. As a result, conclusions drawn on the assumption that everything except the STT change is constant are potentially biased. Second, it is hard to decouple transaction volume into stabilizing and destabilizing components. Thus, it is difficult to decipher which portion of the volume—noise or other—is more affected by the tax. Finally, it is difficult to distinguish between the various ways that STTs can affect security prices. These include adjustments in expectations about the impact of the taxes as well as changes in market liquidity. As a result, the general findings in the empirical literature are inconclusive.

66.4 Key Issues

There are several difficulties with the use, design and effectiveness of implementing STTs. If an STT is applied to one asset class, volume may migrate to an asset class without an STT. Habermeier and Kirilenko (2001) develop a model that shows how a stock transaction tax could decrease demand for stocks and increase demand for derivatives. Matheson (2011) argues that “taxing securities without taxing their derivatives could result in migration of trade from the spot markets to the derivatives markets, with an accompanying increase in leverage and risk.” Therefore, in implementing an STT, the relationship between taxed and untaxed securities should be considered.

While taxing all financial instruments may appear to be the solution, this raises new obstacles. Taxing exchange-traded instruments would be relatively straightforward, but taxing over-the-counter transactions would be both expensive and difficult to implement. Therefore, the types of financial instruments that should be taxed needs considerations.

Further, evidence shows that if an STT is implemented in one jurisdiction and not another, volume will migrate. Pomeranets and Weaver (2011) and Umlauf (1993) show that an STT results in a decline in volume because traders either move their activities to regions without an STT or stop trading entirely. The loss in trading volume may hinder the information efficiency of the financial market because volume contains informational content.

Furthermore, if a large portion of the trading volume shifts to a tax haven as a result of an STT, there will be less revenue collected than originally projected. Since advocates of STTs propose that the tax will raise substantial revenues, the transfer of volume will reduce the number of taxable transactions thereby reducing the potential revenues.

An additional concern with imposing an STT is territoriality. An STT can be applied to transactions based on the location of the trade, the location of the investor, or the location of the securities' issuer. Great care should be taken in defining territoriality policies and definitions as gaps may arise that result in tax evasion. Furthermore, cross-jurisdictional coordination needs to be developed to avoid double taxation. If one country uses residency criteria to tax transactions and another country uses the place of transaction, investors may be at risk of double taxation.

Introducing an STT warrants certain exemptions and limitations. Initial public offerings and transactions in the primary markets should be exempt from an STT. In addition, a limit should be imposed on the maximum tax liability for investors. This limit could be based on transaction volume or size and would reduce the tax burden for individual investors so that large investors would not be at a disadvantage.

If an STT is not appropriately designed it could interfere with the smooth functioning of financial markets, lead to informational inefficiency, arbitrage, tax evasion and double taxation. Therefore, effective implementation of STTs requires cross-jurisdictional coordination, controls on cross-border transactions and carefully constructed enforcement.

66.5 Conclusion

This chapter reviews the theoretical and empirical literature that studies the impact an STT has on financial markets, as well as some key issues associated with STTs. The theoretical arguments on the impact STTs have on market quality are mixed. Some theories predict a negative relationship between an STT and volatility and liquidity, and others point to a positive relationship. As a result, there have been numerous attempts to sort out the debate empirically. Generally, the empirical studies find that an STT is inversely related to volume and directly related to spreads. The literature also finds no significant relationship and in some cases a positive relationship between STTs and volatility. These findings are in stark contrast to what advocates of the tax propose since an STT can increase volatility, and reduce liquidity and the efficiency in financial markets. As a result, when implementing an STT, policymakers will need to consider ways to mitigate the unintended consequences of STTs.

References

- Amihud, Y., and H. Mendelson. 1992. Transaction taxes and stock values. In *Modernizing U.S. securities regulations: Economic and legal perspectives*, ed. Robert W. Kamphuis Jr., 477–500. New York: Irwin Professional Publishing.

- . 2003. *Effects of a New York State stock transaction tax*. Unpublished manuscript, November.
- Baltagi, B., D. Li, and Q. Li. 2006. Transaction tax and stock market behavior: Evidence from an emerging market. *Empirical Economics* 31: 393–408.
- Bloomfield, R., M. O'Hara, and G. Saar. 2009. How noise trading affects market: An experimental analysis. *Review of Financial Studies* 22 (6): 2275–2302.
- Campbell, J., and K. Froot. 1993. *International experience with securities transaction taxes*. NBER working paper 4587.
- Cortez, B.S., and T. Vogel. 2011. A financial transaction tax for Europe. *EC Tax Review* 1: 16–29.
- Green, C., P. Maggioni, and V. Murinde. 2000. Regulatory lessons for emerging stock markets from a century of evidence on transactions costs and share price volatility in the London Stock Exchange. *Journal of Banking and Finance* 24: 577–601.
- Habermeier, K., and A. Kirilenko. 2001. In *Securities transaction taxes and financial markets in taxation of financial intermediation*, ed. P. Honohan, 325–344. New York: Oxford University Press.
- Hau, H. 2006. The role of transaction costs for financial volatility: Evidence from the Paris Bourse. *Journal of the European Economic Association* 4 (4): 862–890.
- Hu, S. 1998. The effects of the stock transaction tax on the stock market – Experience from Asian markets. *Pacific-Basin Finance Journal* 6: 347–364.
- Jones, C.M., and P. Seguin. 1997. Transaction costs and price volatility: Evidence from commission deregulation. *American Economic Review* 87: 728–737.
- Keynes, J.M. 1936. *The general theory of employment, interest and money*. London: MacMillan.
- Kupiec, P. 1996. Noise traders, excess volatility, and a securities transaction tax. *Journal of Financial Services Research* 10: 115–129.
- Matheson, T. 2011. *Taxing financial transactions: Issues and evidence*. International Monetary Fund (IMF) working paper 11/54.
- Pomeranets, A., and D. Weaver. 2011. *Security transaction taxes and market quality*. Bank of Canada working paper 2011–26.
- Roll, R. 1989. Price volatility, international market links, and their implications for regulatory policies. *Journal of Financial Services Research* 3: 211–246.
- Sahu, D. 2008. Does securities transaction tax distort market microstructure? Evidence from Indian stock market. Available at SSRN: <http://ssrn.com/abstract=1088348>
- Saporta, V., and K. Kan. 1997. *The effects of stamp duty on the level and volatility of UK equity prices*. Bank of England working paper no. 71, July.
- Schwert, G.W., and P. Seguin. 1993. Securities transaction taxes: An overview of costs, benefits and unresolved questions. *Financial Analysts Journal* 49: 27–35.
- Stiglitz, J.E. 1989. Using tax policy to curb speculative short-term trading. *Journal of Financial Services Research* 3: 101–115.
- Summers, L.H., and V.P. Summers. 1989. When financial markets work too well: A cautious case for a securities transaction tax. *Journal of Financial Services Research* 3: 261–286.
- Tobin, J. 1978. A proposal for international monetary reform. *Eastern Economic Journal* 4: 153–159.
- Umlauf, S.R. 1993. Transaction taxes and the behaviour of the Swedish stock market. *Journal of Financial Economics* 33: 227–240.

Savita A. Sahay

Abstract

This chapter intends to explore transfer pricing and some popular transfer pricing methods. Since transfer prices are used to assess the incomes of the firm's divisions, choice of a transfer pricing policy can influence many important operational and strategic decisions, such as capital and resource allocation, volume and efficiency of production, performance evaluation as well as tax planning.

We use a descriptive method to study and compare the most popular transfer pricing policies used in American industry and find that many transfer pricing policies may lead to sub-optimal decisions and conflicts within the company because of the incentives they create for managers.

For the purpose of optimal decision making in a multi divisional form, Actual cost-based transfer pricing with an additive markup has been shown to dominate an entire class of alternative policies. In multinational corporations, there is an incentive to reduce the overall income tax burden by charging higher prices to units located in countries with high tax rates. This chapter also explores the economic incentives for the use of a particular transfer pricing policy by multinational companies.

Keywords

Comparison · Divisional performance evaluation · Optimal · Popular methods · Tax planning · Transfer pricing

S. A. Sahay (✉)

Department of Accounting and Information Systems, Rutgers Business School, Rutgers University, Piscataway, NJ, USA

e-mail: savita.sahay@business.rutgers.edu

67.1 Introduction

In decentralized firms, a transfer price is usually assessed for internal transactions in which one division of the firm provides an intermediate product or service to another. The transfer price appears in the income statement of the selling division as part of its revenue, and in that of the buying division as part of its cost. Since managers are typically evaluated and compensated based on the reported income of their divisions,¹ the method used for setting transfer prices directly influences the decisions delegated to them. This makes transfer pricing policy (henceforth TP) an important tool in the hands of upper management for coordinating autonomous subordinate divisions. Additionally, the choice of TP policy is an important one for high-level management since TP is related to many other operational and strategic processes within the firm.²

While the choice of a transfer pricing policy is important and contentious, it is a fascinating and curious phenomenon. Activities within an organization are clearly non-market in nature- products and services are not bought and sold as they are in market transactions. Yet, establishing prices for transfers among subunits of an organization has a distinct market flavor. The rationale for transfer prices is that subunit managers, when evaluating decisions, need only focus on how their actions will affect subunit performance without worrying about their impact on companywide performance. In a well functioning transfer pricing system, optimizing subunit performance leads to optimizing the performance of the organization as a whole.

The Transfer Pricing (TP) problem becomes even more complex when a company has multiple divisions or subsidiaries operating in different countries that have different tax rates, tariff and import duties, currency and foreign exchange restrictions. In multinational corporations, there is an incentive to reduce the overall income tax burden by minimizing profits in the higher tax countries and maximizing profits where the tax rates are lower.

Since large multi divisional firms delegate many decisions to the divisional managers, designing a financial control systems that would coordinate the decisions throughout an organization and will guide the behavior of its employees is an important task. To be effective, financial control systems should reflect an organization's strategies and goals. Since an organization's major goal is to earn a satisfactory profit, the control systems put in place must ensure that the goals of the organization are consistent with the goals of its individual members (the goal congruence objective). Managers and other employees of the company are individuals who might have personal goals, often in contrast with organizational goal. For

¹ According to Merchant (1989, about 95% of the US managers' compensation is based on their divisional income.

² These include tax planning, capital and other resource allocation, costing, decentralized co-ordination and control as well as performance evaluation, and formulation of integration and diversification policy.

example, a company would prefer an employee to exert maximum effort towards achieving a goal, while the employee might prefer to work as little as necessary to keep his job. Management control systems must motivate managers and other employees to exert maximum effort through a variety of monetary (cash, company shares, vacations) and nonmonetary (promotion, pride in working for a successful company) rewards.

In this chapter, we will try to understand how a corporation's senior executives design and implement the financial control systems that are used to plan and control the firm's performance. Although there are many aspects of financial control, such as strategic planning, budgeting, performance measurement, responsibility center allocation and transfer pricing, we will restrict our attention to the two most important aspects of financial control:

1. Design of responsibility centers
2. Selection of an appropriate transfer pricing methodology.

67.2 Designing Responsibility Centers

A desirable feature of financial control systems is that they should be designed to reflect the decision-making responsibility of individual managers. This necessitates assigning financial responsibility (in terms of costs, revenues, profits and assets) to organization subunits. These subunits are called **Responsibility Centers (RC)**. A responsibility center is an organization subunit that is headed by a manager who is responsible for its activities. Since the managers' compensation is often dependent on the financial performance of the division, managers are automatically motivated to maximize the difference between its inputs and outputs. Since a company is a collection of responsibility centers, if each center earns a satisfactory profit, the company as a whole will earn satisfactory profit, a major objective of any profit-oriented organization. In this sense, Financial Control measures how well a company or a RC or any department controls its cost and maximizes its profits.

The core operation of any RC involves receiving inputs in the form of material, labor and services, and transforming them into outputs, either tangible (i.e. goods) or intangible (i.e. services). The products produced by a RC may be transferred either to another RC, where they are inputs, or to the outside marketplace, where they are outputs of the organization, and earn a revenue. To measure their performance, the company may use one of the four types of RCs:

67.2.1 Cost Centers

A division is established as a cost center whenever central management can measure the output, knows the cost function and can set the appropriate quantity of output. Cost center managers are given the authority to determine the mix of inputs (Labor, materials and outside services) and they are evaluated based on their efficiency in

applying inputs to produce outputs. Since they do not sell their product or services themselves, they are not evaluated on profits or revenues.

Cost centers are typically manufacturing departments which are asked to produce a quantity level determined by top management and are either responsible for minimizing cost for a given output or to maximize output for a given budget.

Since the cost center manager is typically evaluated based on minimizing costs, he has the incentive to lower the costs by lowering quality of inputs resulting in a possible reduction in the quality of the final part or output. Therefore the quality of products manufactured in a cost center must be carefully monitored.

In cases where cost center managers are allowed to choose the level of output, they have the incentive to choose the output level at which the center's average costs are minimized, which may not be the optimal production level (Usually the level at which marginal costs are minimized) for the company's profits. Moreover, cost center managers have an incentive to over-report costs so that they can either pocket the slack or exert minimal effort in reducing costs.

Thus, cost centers are ideal for situations where the central management determines the output level, has a good idea of the division's cost function, can observe the quality of the output and can set the appropriate rewards.

67.2.2 Revenue Centers

In a revenue center, output is measured in monetary terms but the expenses are typically not matched with revenue. Revenue centers are usually marketing divisions that do not have the authority to set selling prices and are evaluated against target revenues or budgets. A center that is responsible for both expenses and revenues is called a profit center.

67.2.3 Profit Centers

Profit centers are divisions responsible for both the revenues and the costs and are typically evaluated based on their efficiency in applying inputs to produce outputs. Profit centers are set up when the knowledge required to make the product is firm specific, the internal decisions regarding quantity, quality and pricing of the output are proprietary information, costly to transfer outside the division.

Since profit center managers are responsible for maximizing the division's profits, rewarding them with a portion of their divisional income seems straightforward. However, there are a large number of profit centers that produce an intermediate product, to be transferred within the company. As we will see later, companies often use a Transfer Price to evaluate these transactions. The choice of an inappropriate transfer pricing method can lead to suboptimal quantity, quality or input mix decisions by the profit center manager.

A second problem arises when company's overhead costs are to be allocated to these business units. The allocation of corporate overhead costs is often arbitrary and is under constant debate because it reduces the total profit of the center.

As a result, profit maximization by an individual profit center may not maximize profits for the firm as a whole due to inter dependencies among different centers. For example, a profit center using resources to advertise its product might hurt the sales of another product by the same company, leading to overall lesser profit for the firm as a whole.

To help managers internalize the effect of their decisions on the company as a whole and other profit centers, profit center managers are often compensated based on their own as well as firm wide profits.

67.2.4 Investment Centers

Investment centers are similar to profit centers except that they have additional authority to take capital investment decisions. They are typically evaluated on measures such as return on investment or residual income. Investment centers are established when the manager has specific knowledge about investment opportunities and possesses information specific to the operating decisions. An investment center may have many profit centers and cost centers inside it.

67.2.4.1 Return on Investment (ROI)

Return on Investment (ROI) is used by a large majority of US firms as a method of compensating investment center managers. It is the ratio of operating net income divided by the total investment made in the center. The method has intuitive appeal because it compares the center's performance automatically with other centers and with other industry benchmarks.

However, ROI creates incentives for sub-optimal decision-making by the center manager, hurting the company's overall profit or quality or reputation or long term prosperity. For example, the manager may try to increase operating income by decreasing depreciation, research and development and other discretionary expenses. He may try to artificially increase the revenue (Hence inflate the income, the numerator in the equation) by a practice known as Channel Stuffing, in which managers will supply the product around the year-end to customers, even though they neither ordered nor need them.³ This action may record additional revenue at the year end for the investment center, but it will lead to returned goods costing money and reputation.

³See "Blind Ambition" in Business Week, Oct. 23, 1995, pp. 78–92. According to Business Week article on Bausch and Lomb "Under pressure to beat sales target in 1993, contact lens managers shipped products that doctors never ordered while assuring them that they wouldn't have to pay until they sold the lenses."

Additionally, total value of investments used in the denominator of the ROI calculation is also subject to gaming behavior by the center managers. Since Accounting rules dictate that companies anticipate all losses but cannot record any gain in market value, the market value of the asset value in the denominator is not really a correct measure of the center's investment. In fact, managers have been known to sell company's fleet of cars and other assets in the last quarter to keep the denominator as small as possible.

Finally, use of ROI as the sole method for performance evaluation can lead to other dysfunctional decisions by the center managers. For example, Investment center managers have incentives to reject profitable projects whose ROI are below the current average ROI for the division because accepting any such project would lower the overall ROI of the division. If the firm is in an industry where the initial projects earns a really high ROI, then managers will continue to search for such high-return projects; rejecting everything meanwhile.

Another disincentive created by ROI is the managers' reluctance to take high-risk projects. Since high-risk projects have a higher cost of capital and also a higher probability of failing, managers would not like any additional risk imposed on them. Also, a manager with a short-term horizon may pick projects that boost ROI in the short run even though they incur huge long-term costs.

67.2.4.2 Residual Income (RI)

To address some of the concerns related to ROI, some companies use residual income to evaluate performance. **Residual income (RI)** measures the "abnormal returns" earned by the division. Therefore a normal rate of return, measured either by the cost of capital or a required rate of return for the center. Is subtracted from division's profits. Under the residual income approach, manager has incentive to select the project as long as the residual income is positive, which simply means that the divisional profit rates are higher than the required rate of return.

However, RI has some problems of its own. Comparing RIs across divisions is difficult since larger divisions are more likely to have higher RI than smaller divisions, since RI is an absolute number.

Moreover, required rate of return is likely to be different for each division, depending on the riskiness and complexity of the project, making it to be a popular source of conflict in firms. Finally, RI suffers from the same problems as ROI, when it comes to the managers with short term horizon- they may try to increase short term RI by cutting on expenses like maintenance, jeopardizing future firm value and cash flows.

67.3 Transfer Pricing

Transfer price is a price paid by one division of a firm to another division for some goods and services they may exchange. The transfer price appears as revenue in the books of the supplying division and as cost in the books of the buying division. Since divisional managers care about their divisional income, choice of a transfer pricing

policy can be an important tool in the hands of the upper management to control and influence the decisions of the managers.⁴ As a general rule, the transfer price policy should be designed to accomplish the following objectives:

1. It should induce goal congruent decisions- that is, the system should be designed so that decisions that improve the division's profits will also improve company profits.
2. It should help measure the economic performance of the division.
3. It should provide each division with the information it needs to take better decisions.
4. It should be simple to understand and easy to implement.
5. Tax considerations and other objectives in International markets.

The variety of TP policies in common use⁵ in American industry suggests that no single TP policy can be expected to be optimal for all firms. The structure of the firm and features of its environment are likely to determine if one policy is better for it than others.

The fundamental issue is that the transfer price should be similar to the price that would be charged if the product were sold to outsiders or purchased from outsiders, correcting for the savings in marketing and other transaction costs. The rule is complicated to use in situations when no outside market exists for the exact same goods (For example, an engine produced for a sports car of a specific brand) or the product is hard to specify in the beginning (For example, a large computer program).

67.3.1 Transfer Pricing Methods

There are three methods commonly used to determine transfer prices.⁶ We will show how the choice of transfer pricing method can affect the companywide operating income as well as lead to suboptimal decision making by divisional managers-Suboptimal in the sense that they may maximize the division's income but may hurt the overall company profits.

Transfer prices can vary from very simple to very complex, depending on the nature of the product. Let's start with the simplest first:

⁴For a detailed literature review of transfer pricing as a coordinating mechanism, see Adams and Drtina (2008)

⁵In Tang's (1992) survey of Fortune 500 companies, there is evidence of companies using negotiated transfer pricing, market based transfer pricing as well as several variations of cost based transfer pricing.

⁶See for example, Kaplan (1982) for one of the first detailed discussion of various TP methods.

67.3.1.1 Market Based Transfer Prices

Since market based transfer price will best simulate a competitive environment within the firm, a market based transfer price is best to use when there is a market price available reflecting the same quantity, quality and other features as the product being transferred. The argument is that if the firm cannot make a profit at the market price, then the company is better off not producing it and purchasing it from outside. By using a market price in a competitive market, a company can not only take the correct make or buy decisions, it can also achieve almost all the objectives of a transfer pricing policy: goal congruence, measurement of manager's effort, division's autonomy, simplicity as well as performance evaluation. It is also likely to lead to minimum conflicts since most parties will consider market-based price to be fair.

A market price for a similar product is often used in the absence of an identical product. Conflict is likely to set in when the adjustment to the external price is made to reflect the true opportunity cost of the product to the firm. There are several reasons for this conflict. Firstly, internal transactions are usually cheaper than outsourcing due to synergies or interdependencies among products, reduced transportation costs, reduced marketing and administrative costs and reduced costs of bad debts. Making an adjustment to the market price to reflect all these savings is likely to be arbitrary. Secondly, even if the monetary costs are the same, firms are more likely to prefer to produce internally due to greater quality control, timely supply or greater control of proprietary information. When all these factors are included in the adjustment to the external price, the resulting transfer price may not seem fair to all.

Therefore, the ideal situation to implement a market based transfer price would be one where the managers are free to buy and sell in the outside market if it is in their best interest. If buyers cannot get a satisfactory price from the inside source, they are free to buy from outside. Similarly, if the seller can get a better price by selling it outside, it should be allowed to do so. In these circumstances, transfer price will represent exactly the opportunity cost to the seller of selling the product inside and lead to goal congruent decisions.

67.3.1.2 Cost Based Transfer Prices

If there is no external market for the intermediate good, then a transfer price based on some measure of cost may be considered:

67.3.1.3 Variable Cost Transfer Prices

Since the goal is to set a transfer price that reflects the true opportunity cost to the firm, variable production cost may be the most effective transfer price.⁷ Variable costs also receive support from economists since they are considered to be closest to marginal costs, the method of choice for Economists.⁸ The case for marginal cost is

⁷For a theoretical model of Variable Cost based Transfer Pricing, see Vaysman (1996)

⁸Hirshleifer (1956) established conditions under which pricing internal transfers at marginal cost maximizes firm profit.

that that if the transfer price is higher than the marginal cost, the supplying division would like to sell more than the optimal quantity and the buying division would like to buy less than the optimal quantity, leading to suboptimal production decisions.

However, transfer pricing at marginal cost has been criticized by some experts. Eccles (1985, 22) notes that transfer pricing without regard to fixed costs, overhead and profit for the selling division leads to an unfair measure of its contribution to the company. This criticism addresses the fact that the manufacturing division does not recover its fixed costs. Thus, the division would appear to be losing money.

A related problem with variable cost transfer pricing is that it creates incentives for the manufacturing division to distort variable cost upward, perhaps by misstating some fixed costs as variable costs. Since cost classifications for many costs are to some extent arbitrary, serious conflicts within the firm can arise.

Lastly, manufacturing division has the incentive to try to convert some of its fixed costs into variable costs, simply to get reimbursement, even though this conversion may be costly to the firm as a whole. For example, the division might choose to hire a costly consultant instead of hiring an employee or lease an asset instead of buying it to keep variable costs high and fixed costs low, since it gets reimbursed for the variable costs but not for the fixed costs.

There are several variations of variable costs to consider as transfer prices:

67.3.1.4 Standard Variable Cost Transfer Prices

A standard cost is a carefully determined cost of a unit of output. It may also represent the future cost of a product, process or subcomponent. Standard costs usually arise from the budgeting process of the company and represent the expected cost that will be used as a benchmark to compare with actual operating costs for performance evaluation purposes.

Standard costs are useful for both decision-making and control. Since, standard costs are known to all well before the actual variable costs become known, division managers have incentives to keep actual costs down, within the standards prescribed or their performance will suffer. Also using standard costs as benchmark allows top management to use deviations between actual costs and standard cost (Variance analysis) effectively for cost control and improvement in future performance. Lastly, since actual costs may vary from one batch to another, using standard costs are easier on the accounting system.

However, standard costs have been criticized for many reasons. First, depending on the philosophy of top management, standard costs may mean something that is currently achievable or something that can be achieved with abnormal effort or luck. Too tight a standard might discourage the supplying division, while too loose a standard might not encourage maximum possible effort.

Second, most firms are reluctant to change standards during the year and often thereafter, unless a large, unexpected change occurs in a standard. Knowing the relative rigidity of the standards encourages suboptimal strategic behavior from the divisional manager. For example, manufacturing division managers, who are often involved in setting of standards, may overstate costs and keep a “slack” which they can consume, in form of lower effort, throughout the year. On the other hand, if

standards are revised frequently during the year, divisional managers have less incentive to control costs. Therefore the decision to revise standards often involves the trade-off between loose standards and tight standards. Each of them can be a source of conflict between trading parties.

67.3.1.5 Actual Variable Cost Transfer Prices

A common method for arriving at the transfer price is to use the actual cost of production of the transferred product. A transfer pricing policy based on actual cost is usually easy to implement. It is also attractive from an internal accounting perspective since it generates an income statement that eliminates intra-company profits.

It is usually said that transfer prices based on actual costs provide no incentive to the supplying division to control costs, since the supplier is reimbursed for his costs. Also, since actual costs may vary from one batch to another, keeping track of batch costs and inventory may be hard for the accounting system. Another problem with actual cost based approach can arise in firms where operations are capacity constrained. When a firm is operating at capacity, production decision should optimize the use of the capacity rather than individual product's profit margin. In such cases, actual cost base transfer price will lead to a suboptimal quantity decision. In firms operating at full capacity, transfer price should instead be the sum of marginal costs and the opportunity cost of capacity, which would be the profit of the best alternative use of the capacity.

Because of the incentive and information problems described above, a commonly used variation of variable cost transfer pricing is to price all transfers at a **cost plus mark-up**. The mark-up may be in the form of a fixed fee to cover manufacturing division's fixed cost. Some firms choose a fixed fee to cover fixed cost plus a return on equity.

The most well known example of a markup is the *multiplicative* markup, in which the unit transfer price is determined by multiplying a cost measure by a fixed factor. For example, it might be set equal to 120% of the cost. Many accounting textbooks⁹ use the multiplicative method to illustrate the character of cost-plus transfer pricing.

An alternative to the multiplicative method is the *additive markup*, in which the unit transfer price is set at a fixed dollar amount over cost (e.g., the variable cost plus \$5.) Evidence for the use of such methods appears in a survey of Fortune 500 companies conducted by Price Waterhouse. The survey revealed that about 45% of all companies use an approach that "does not rely on percentage calculations" (Price Waterhouse, 1984, 13). Instead, "the cost and markup are maintained as two pieces of information and intra-firm profit is determined by adding up the markups."

Efforts to compare different transfer pricing methods show that additive markup is optimal among the class of cost-plus methods.¹⁰ The argument is that additive

⁹For example, Garrison and Noreen (2012)

¹⁰See Sahay (1997) and (2003).

markups are not only convenient to keep separate track of costs and profits, they also motivate the manufacturing division to reduce the cost so that it can maximize the number of units transferred (Hence the amount of mark-up earned). In contrast, under a multiplicative mark-up, the manufacturing division's income (Mark-up) is proportional to the production cost. Hence, the division is more interested in increasing the cost, and the resulting mark-up, than transferring the optimal quantity.

67.3.1.6 Full Cost Transfer Prices

Since most transfer pricing methods discussed so far suffer from some incentive problem or accounting complication, one alternative is to use a simple and objective transfer-pricing rule that will avoid wasteful disputes and arbitration costs. Since full cost is the sum of fixed and variable cost, it is easily available in the accounting system and cannot be changed by reclassifying a fixed cost as variable cost.

Another reason for the popularity of full cost transfer pricing is its ability to deal with the problem of changes in capacity. As a plant begins to reach capacity, opportunity cost to produce one more unit is likely to rise higher than variable costs. In such cases, full cost might be a closer approximation to opportunity cost than variable cost.

The incentive problem that most cost based systems suffer from is also present here. They provide no incentive to the manufacturing division to control costs, since the costs can always be recovered via transfer price. In fact, full cost method allows manufacturing division to transfer all of its inefficiencies to the buying division.

From optimality standpoint, the problem with full cost transfer pricing is that the quantity traded will always be suboptimal. Since full cost overstates the opportunity cost of producing and transferring one more unit internally, manufacturing division would like to produce and transfer more units while the buying division would like to buy too few a units.

Despite all of these problems, full cost transfer pricing appears to be very common in practice. Perhaps the reason for its popularity is its simplicity and low cost of implementation. Scant theoretical research also suggests that with an appropriately chosen markup, the performance of actual cost based transfer pricing can be significantly improved, even though the markup leads to suboptimal resource allocation decision within the firm.¹¹ This result is consistent with the conventional wisdom of the accounting profession that judicious use of transfer pricing in a decentralized firm must trade off pricing efficiency with effective incentives for divisional decision making that would further the long term interests of the firm. Empirical evidence also suggests that firms that use cost based transfer pricing are well aware of its incentive problems, a vast majority of them choosing to price transfers above incremental production costs (see Price Waterhouse, 1984 survey and Tang 1992 survey).

¹¹ See Sahay (2003).

67.3.1.7 Negotiated Transfer Prices

Several surveys of transfer pricing practices among large, decentralized firms document the use of negotiated transfer pricing policies. Such a policy lets divisional managers negotiate all aspects of an intra-firm transaction, including its price. Since divisional managers have superior local information, minimum interference from upper management in the trading decision enhances the 'flexibility gain' associated with decentralization. In particular, if the divisions have symmetric information, negotiated transfer pricing is likely to lead to efficient intra-firm trade: the divisions will choose to maximize the joint contribution margin, negotiations being limited to how this margin will be shared.

While negotiation is a fairly common method, it too has drawbacks. Firstly, divisional performance becomes sensitive to the relative negotiating skills of the two divisional managers. Moreover, if the supplying division manages to negotiate a price higher than the optimal price, implying that the receiving division will get lower than the optimal price, the quantity traded will be less than optimal, leading to lower firm-wide profits.

Secondly, negotiated transfer pricing can produce conflicts among divisions, leading to time and resource consuming arbitrations.

Finally, It is well known that negotiated transfer pricing suffers from under-investment due to the 'hold-up' problem. This problem has received much attention in the literature on incomplete contracts (see, for example, Williamson 1985) and occurs when the parties to a trade have the opportunity to make specific investments before the date on which they negotiate a price for the traded good. In such a setting, the investment costs must be considered 'sunk' when negotiating the price since they would have to be borne even if negotiations were to fail. Thus, the parties will be reluctant to incur any costs *ex ante*, and investments will be 'held up.'

Edlin and Reichelstein (1995) show that the hold-up problem associated with negotiated transfer pricing may be eliminated if the parties are allowed to sign a contract before the investments are chosen.¹² In essence, this prior contract (which serves as the *status quo*, should subsequent negotiations fail) provides protection for the costly investments. However, as argued by Holmstrom and Tirole (1991) inter-divisional transactions in a large firm are governed by incomplete contracts that are impossible to enforce, making the hold-up problem unavoidable.¹³

The discussion above suggests that no single TP policy can be expected to be optimal for all firms. The structure of the firm and features of its environment are likely to determine if one policy is better for it than others. It is somewhat surprising, therefore, that there has been little theoretical research on the comparison of alternative TP practices or on the setting up of general guidelines based on economic principles that can help a firm in its TP policy-making. Part of the reason for this state

¹² Chung (1991) derives a similar efficiency result, also using a prior contract.

¹³ See also the Executive Summary of the Price Waterhouse (1984) survey which reports: "Most companies (72%) state that they do not use formal contracts to document internal buy/sell agreements" (p. ii).

of affairs is that it is extremely difficult to address the TP problem without touching upon a host of other problems to some degree; a theory of TP is really part of a deeper theory of decentralization.¹⁴ Another reason is the very variety of corporate TP practice which makes the problem important and interesting: it is difficult to analyze various methods of administered and negotiated TP¹⁵ in a single theoretical model, in order to extract principles that can be generally applicable.

A few efforts in theoretical research have been made to produce a systematic study of the policies governing intra-firm transactions in a large, multi-divisional firm. The purpose of this research is to produce a systematic study of the policies governing intra-firm transactions in a large, multi-divisional firm. The results point towards emergence of a theory that addresses questions as to which TP policy is best for a given firm. Findings suggest that firms should be cautious in putting a particular TP policy in place, since the choice of the policy could significantly affect firm's profits, volume and efficiency of production and managers' incentives in taking other decisions important to the firm. In general, several economic factors have been identified that make one TP policy preferable to other. These economic factors include revenue and cost characteristics of the intermediate product transferred, the number of products, the price elasticity and the level of competitiveness of the external market faced by the firm.

In one of the first papers (BRS 99), a comparison of negotiated transfer pricing (NTP) and standard cost-based transfer pricing (SCTP) reveals general superiority of NTP. In a second paper (Sahay 2003) I show that the performance of actual cost-based transfer pricing with an additive markup is superior to any other cost-based policy in its class. Specifically, the actual cost-based method is shown to outperform several cost-based transfer pricing schemes, including SCTP. Sahay (2012, Working paper)¹⁶ attempts to complete the ranking, showing the superiority of actual cost-based transfer pricing over negotiated transfer pricing, as long as the buying division's investment is sufficiently important.

67.4 Some Theoretical Models of Transfer Pricing

There is a considerable body of research available on theoretical transfer pricing models. However, only a few of these models are used in actual business situations. Although these models are not directly applicable to the real business situations, they

¹⁴See Holmstrom and Tirole (1991) for an elaborate discussion of this point and for a model that endogenizes the need for transfer pricing in a more general study of possible decentralized forms. In this chapter, however, we take transfer pricing as a control mechanism already in place in the firm.

¹⁵The literature classifies TP policies as *administered* (where most rules are laid down by headquarters) or *negotiated* (where participating divisions formulate their own rules.) Administered TP is either *cost*-based or *market*-based while negotiated TP often uses cost or market price as a starting point for negotiations. See the surveys of Eccles and White (1988) and Price Waterhouse (1984) for more details.

¹⁶For a comparison of actual cost based with NTP, see Sahay, S. (2008)

are useful in conceptualizing transfer pricing systems. These models may broadly be divided into three types:

1. Models based on economic theory,
2. Models based on linear programming, and
3. Models based on Shapley value.

67.4.1 Economic Models

The classic economic model was first described by Hirschleifer¹⁷ as a series of marginal revenue, marginal cost and demand curves for the transfer of an intermediate good from one division to another. He used these curves to establish transfer prices that would optimize the total profit of the two business units.

The difficulty with this model is that it can be used only when a specifies set of conditions exists: It must be possible to estimate the demand curve of the product, there can be no alternative use of facilities used to make the intermediate good and the selling unit makes only one product which it transfers to a single buying unit. Such conditions are rarely met in the real world.

This model (and other economic models) assumes that transfer prices will be dictated by central management. It, therefore, denies the importance and benefits of negotiation among trading units. As a result, using the model may lead to delayed and dysfunctional decision making since negotiation between divisions leads to better decisions regarding price and quantity.

67.4.2 Linear Programming Models

The linear programming model is based on an opportunity cost approach. The model calculates an optimal production pattern for the firm and uses this production pattern to calculate a set of values that impute the profit contributions of each of the scarce resources. These are called *shadow prices* and one way of calculating them is called “obtaining the dual solution” to the linear program. If the variable costs of the intermediate good are added to their shadow prices, a set of transfer prices results that should motivate divisions to take optimal production decision for the firm. This is so because, if these transfer prices are used, each business unit will optimize its profits by producing in accordance with the patterns developed through the linear program.

If reliable shadow prices could be calculated, this model would be useful in arriving at the transfer prices. However, many simplifying assumptions are needed to make the model work. Some of these assumptions are that the demand curve is

¹⁷Jack Hirschleifer. On the economics of transfer pricing. *Journal of Business*, July 1956, pp. 172–184

known, that it is static, that the cost function is linear and that the incremental profits from alternative use of factory can be estimated in advance. It is hard to meet these assumptions in the real world.

67.4.3 Shapely Value

Shapely value was developed in 1953 by L. S. Shapley as a method of dividing the profits of a coalition of individuals in proportion to each individual's contribution. It is used commonly in the theory of games as the most equitable solution to this problem.

Whether the same technique can be applied in transfer pricing is a debatable issue. Although the technique has been around for many years, few practical applications have been reported. A possible reason for its lack of popularity could be the complexity of computation unless there are only a few products involved in the transfer. Another reason is the general belief that the Shapley method is not valid for solving the transfer-pricing problem.

67.5 Optimal Transfer Price

Although no general rule always meets the goal of choosing the best transfer pricing policy, some guidelines¹⁸ and boundaries can be established:

67.5.1 Minimum Transfer Price

The minimum transfer price acceptable to the selling division is clearly the variable unit cost of the product. Since, fixed costs are considered to be sunk costs, selling division would be interested in trading as long as its out of pocket costs are covered.

This is true, however, if the selling division has sufficient capacity to produce the entire order and would not have to give up some of its regular sales. In cases where capacity constraints force the division to either transfer an item internally or sell it externally- that is, it cannot produce enough to do both, then the selling division would expect to be compensated for the contribution margin on those lost sales. In general, if the transfer has no effect on fixed costs, then from the selling division's standpoint, the transfer price must cover both the variable cost of producing the transferred units and any opportunity costs from lost sales.

From the firm's standpoint, transfer is desirable if (1) the total cost of producing the good (by both divisions) is less than the price it can receive for the good in the

¹⁸See general guidelines regarding Transfer Prices in Horngren et al. (2011)

outside market and (2) it does not pay more to produce the good internally than it would have to pay to buy it in the marketplace. The only transfer price that would achieve both these objectives for the firm is the formula suggested below:

$$\text{Minimum Transfer Price} = \text{Seller's variable cost} + \text{opportunity cost}$$

67.5.2 Maximum Transfer Price

From the buying division's perspective, the trade is beneficial only if its profit increases. For that, it must be able to sell the final product for more than the transfer price plus other costs incurred to finish and sell the product. So the maximum transfer price that it can offer is the difference between the final price and additional variable costs incurred by the buying division. This transfers the entire surplus from the transaction to the selling division. For example, if a good can be sold in the market for \$30 and the variable costs of selling and buying division are \$10 and \$6 respectively, then the buying division can pay up to \$24 (\$30 – \$6). Anything more would put him at a loss.

In cases where the buying division has an outside supplier available, the choice of maximum transfer price is simple. Buy from an inside supplier only if the price is less than the price offered by the outside supplier. This may lead to suboptimal decision from the firm's standpoint. For example, if a good can be sold in the market for \$30 and the variable costs of selling and buying division are \$10 and \$6 respectively, then per unit profit for the company is \$14. However, if the transfer price is set at \$12 and an outside supplier is willing to provide it for \$11, the buying division would buy it from outside, even though the company could have spent only \$10 in producing it internally. So, the highest transfer price in this case is \$11, the alternative maximum price from an outside source.

67.5.3 Idle Capacity and Transfer Price

As mentioned before, idle capacity can significantly change the economics and psychology of transfer pricing. If selling division has large unused capacity, more than enough to satisfy the buying division's demand, then it would be interested in the proposal as long as its variable cost is covered. Since there would be no lost sales, there is no opportunity cost, minimum transfer price would be equal to variable cost. (\$10 in the example above).

Maximum transfer price would, once again, depend on the availability of an outside price. If the buying division can buy similar product from an outside vendor for \$15, then it would be unwilling to pay more than \$15 as the transfer price.

Thus combining the requirements of both the selling and the buying division, the acceptable range of transfer price would be between \$10 and \$15.

67.5.4 No Idle Capacity and Transfer Price

Generally, firms prefer internal transactions to external ones. After all, firms are organized as a collection of profit centers due to synergies and savings in transaction, bargaining, marketing and other administrative costs and would prefer to produce a part internally than buy it from outside. Other reasons for firms to prefer an internal transaction may be quality control, timely delivery and security of proprietary information. So if the selling division is selling its entire capacity production to outside market, it would have to divert some product away from its regular customers to be able to fill an order from the buying division.

In such cases, the minimum transfer price would be unit variable cost plus the unit contribution margin from lost sales. To continue the example above, suppose that the selling division is selling its entire capacity of 1,000 units to outside market at \$15 per unit and receives an order of 200 units to be supplied to the internal division. So the minimum price, that the selling division is willing to consider as transfer price, is its unit variable cost (\$10) plus the unit contribution margin on lost sales ($\$5 = \$15 - \$10$).

The maximum transfer price, as before, would be equal to the cost of buying it from an outside supplier. Thus, if the outside vendor is ready to supply the good at \$18 as in the example above, the transfer price would be set between a range of \$15 and \$18.

67.5.5 Some Idle Capacity and Transfer Price

If the selling division has only some idle capacity, but not enough to fill the entire order by the buying division, then it would have to divert only some of the product from its regular customers, keeping the opportunity cost portion of the minimum transfer price at a lower level. In our example, suppose the selling division is currently selling only 900 units when its capacity is 1,000 units, it can supply only 100 units internally without diverting sales from its regular customers. However, the buying division needs 200 units. Let us also assume that the selling division will have to supply the entire order of 200 units, having to divert 100 units from its regular customers. Thus the minimum transfer price would be variable cost ($\$10 \times 200 \text{ units} = \$2,000$) plus the unit contribution margin on lost sales ($\$5 \times 100 \text{ units} = \500). Since the transfer quantity is 200 units, unit transfer price would be $\$12.50 = (\$2,000 + \$500)/200 \text{ units}$.

67.6 International Transfer Pricing

In today's global markets, companies may produce goods and services domestically and sell them internationally or produce them outside the country and sell them here. Since the profit is earned in the country of the sale, differences in tax laws can be the

leading determinant of transfer pricing choices. Tax factors include, not only income taxes but also payroll taxes, custom duties, tariffs, sales taxes, environment related taxes and other government levies on organizations. Lax tax laws in one country can encourage a Multi-National Corporation to deploy resources in that country. Choice of an appropriate transfer pricing policy can help in minimizing a company's tax burden, foreign exchange risks and can lead to better competitive position and governmental relations. Although domestic objectives such as divisional autonomy and managerial motivation are always important, they often become secondary when international transfers are involved. Companies would typically focus on charging a transfer price that would reduce its tax bill or that will strengthen a foreign subsidiary.

For example, a company may choose a low transfer price for parts shipped to a foreign subsidiary to reduce custom duty payments or to help the subsidiary to compete in foreign markets by keeping the subsidiary's costs low. On the other hand, it may choose to charge a higher transfer price to draw profits out of a country that has high income tax rates to a country that has lower tax rates or out of a country that has stringent control on foreign remittances.

Transfer pricing is a major concern for multinational companies as highlighted by the fact that approximately 80% of Fortune 1,000 firms select transfer pricing policies keeping financial, legal and other operational considerations in mind. In addition, intra-firm trade accounts for about 55% of the trade between Japan and EU, and 80% of the trade between US and Japan.

Tax authorities are aware of the incentives to set transfer prices to minimize taxes and import duties. Therefore, U.S. Internal Revenue Service (IRS) pays close attention to taxes paid by multinational companies within their boundaries. At the heart of the issue are the transfer prices that companies use to transfer products from one country to another. For example, in 2004, the IRS fined U.K. based pharmaceutical manufacturer GlaxoSmithKline \$5.5 billion in back taxes and interest, stemming from a transfer pricing dispute regarding profits from 1989 to 1996.

It is not surprising, therefore, that most countries have restrictions on allowable transfer prices. Section 482 of the U.S. Internal Revenue Code governs taxation of multinational transfer pricing. This section requires that transfer prices between a company and its foreign subsidiary, for both tangible and intangible property, equal the price that would be charged by an unrelated third party in a comparable transaction. In other words, transfer prices must reflect *an arm's length price*. For tangible goods, the best method is considered to be *a comparable controlled price* between unrelated firms. While this method is theoretically sound, it is difficult to use in practice because intra firm transactions are hard to compare with open market transactions.

The next best method is the *resale price approach*, which is allowed if there is no comparable price available. The method allows all costs and a reasonable profit to be deducted from the resale value.

The third method is the *cost plus approach*, where an appropriate mark up is added to the manufacturing cost to determine the transfer price.

In summary, companies are allowed to use a market based or a cost plus method as long as the surplus represents margin on comparable transactions. In addition, international tax treaties regulate allowable transfer pricing methods.

Eun and Resnick (2012) discuss a case study where the firm's choice of a transfer pricing policy with low markup on cost versus a transfer pricing policy with a high markup result in different incomes and tax burdens for the companies operating in countries with different tax rates, exchange restrictions, import duties and other regulations affecting transfer prices.

67.7 Limitations of Financial Control Systems

An important goal of a business enterprise is to optimize shareholder returns. However, optimizing short-term profitability does not necessarily ensure optimum shareholder returns in the long run. At the same time, the need for ongoing feedback and performance evaluation requires companies to measure a division's performance in the short term, usually at least once a year. Overemphasis on financial controls may be dysfunctional for this as well as many other reasons:

- Firstly, it may encourage short-term actions that are not in the company's long-term interests. For example, the manager may use inferior raw materials to keep costs down, adversely affecting company's goodwill and future sales.
- Secondly, divisional managers may not undertake useful long-term actions, in order to maximize short-term profits. For example, managers may not invest sufficiently in research and development' since R&D investments must be expensed in the year in which they are incurred but their benefits show up only in the future.
- Thirdly, using short-term profit maximization as the sole criteria for performance evaluation can lead to gaming behavior by managers. For example, divisions may set profit targets that can be achieved very easily so that they are never penalized for missing the target. This would result in suboptimal quantity produced and sold. Leading to lower profit for the firm.¹⁹
- Lastly, tight financial control may motivate managers to manipulate data to meet current profit targets. For example, managers may make inadequate provision for bad debts and warranty claims. Managers may also be tempted to falsify data-that is, deliberately provide inaccurate information.²⁰

¹⁹ Another example of tactics used by manager is found in Bausch and Lomb where unusually long credit terms were extended to customers in exchange for big orders. See "Blind Ambition" in Business Week, Oct. 23, 1995, pp. 78–92.

²⁰ According to Business Week article on Bausch and Lomb "Under pressure to beat sales target in 1993, contact lens managers shipped products that doctors never ordered while assuring them that they wouldn't have to pay until they sold the lenses."

In short, relying on financial control measures alone is insufficient to ensure long-term profit maximization. The solution is to use a blend of financial (costs, revenue, profits) and nonfinancial measures (quality, customer satisfaction, innovation) that would achieve the long-term success of the organization.

67.8 Conclusion

This study analyzes the performance of various transfer pricing policies used by American industries and compares them with regard to firm profits and efficient decision making by divisional managers.

Since most firms use transfer prices to assess internal transactions, choice of a transfer pricing policy affects the reported income of the divisions. Since managers are typically evaluated and compensated based on the reported income of their divisions, the method used for setting transfer prices directly influences the decisions delegated to them. This makes transfer pricing policy an important tool in the hands of upper management for motivating managers to take decisions that are in the best interest of firm's overall profits.

The variety of TP policies in common use suggests that no single TP policy can be expected to be optimal for all firms. This study compares the performance of popular TP policies in a large multi-divisional firm. We seek to identify economic factors that make one TP policy preferable to other. These economic factors may be revenue and cost characteristics of the good transferred, the number of products, the price elasticity and the level of competitiveness of the external market faced by the firm.

The findings of the study suggest that firms should be cautious in putting a particular TP policy in place since the choice can be crucial to firms' performance. It seems that among the class of cost based TP policies, an additive markup over variable cost is most desirable from the optimal resource allocation standpoint because it motivates the seller to reduce costs and increase production.

References

- Adams, L., and R. Drtina. 2008. Transfer pricing for aligning divisional and corporate decisions. *Business Horizons* 51–5: 411–417.
- Baldenius, T., S. Reichelstein, and S. Sahay. 1999. Negotiated versus Cost based transfer pricing. *Review of Accounting Studies* 4: 67–91.
- Chung, T. 1991. Incomplete contracts, specific investments and risk sharing. *Review of Economic Studies* 58: 1031–1042.
- Eccles, R. 1985. *The transfer pricing problem: A theory for practice*. Lexington: Lexington Books.
- Eccles, R., and H. White. 1988. Price and authority in inter-profit center transactions. *American Journal of Sociology* 94 (Supplement): S17–S48.
- Edlin, A., and Reichelstein, S. 1995. Specific investment under negotiated transfer pricing: An efficiency result. *The Accounting Review* 70–72.
- Eun, C., and B. Resnick. 2012. *International financial management*. 6th ed. McGraw Hill.
- Garrison, R.H., and E.W. Noreen. 2012. *Managerial accounting*. 14th ed. McGraw Hill.

- Hirshleifer, J. 1956. On the Economics of Transfer Pricing. *Journal of Business*, July 1956.
- Holmstrom, B., and J. Tirole. 1991. Transfer pricing and organizational form. *Journal of Law, Economics, and Organization* 7: 201–228.
- Horngren, C., G. Foster, and S. Datar. 2011. *Cost accounting: A managerial emphasis*. 13th ed. Englewood Cliffs: Prentice-Hall.
- Kaplan, R. 1982. *Advanced management accounting*. Englewood Cliffs: Prentice-Hall.
- Merchant, K. 1989. *Rewarding results: Motivating profit center managers*. Harvard Business School Press.
- Price Waterhouse. 1984. *Transfer pricing practices of American Industry*. New York.
- Sahay, S. A. 1997. Studies in the theory of transfer pricing. Dissertation, UC Berkeley.
- Sahay, S.A. 2003. Transfer pricing based on actual cost. *Journal of Management Accounting Research* 15: 177–192.
- Sahay, S. 2012. Negotiated vs. Actual transfer pricing. Working paper, Rutgers University.
- “Transfer Pricing 2003 Global Survey.” *International Tax Journal* 30(3): 1–86.
- Tang, R. 1992. Transfer pricing in the 1990s. *Management Accounting* 73–8: 22–26.
- Vaysman, I. 1996. A model of cost-based transfer pricing. *Review of Accounting Studies* 1: 73–108.
- Williamson, O. 1985. *The economic Institutions of Capitalism*. New York: The Free Press.



Alternative Models for Evaluating Convertible Bond: Review and Integration

68

Cheng-Few Lee, Lie-Jane Kao, and Po-Cheng Wu

Abstract

This work reviews the literature on convertible debt valuation with various assumptions regarding the firm's market value (or stock price) and interest rate (or straight bond's price). For deterministic stock price and interest rate, the graphical method by Brigham (J Finance 21:35–54, 1966) is a representative. For stochastic variables, the valuation model can be further divided into static and dynamic depending on the valuation is for some known date or for some period of time in the future: the calculus models proposed by Baumol, Malkiel, and Quandt (Q J Econ 80:48–60, 1960), Poensgen (Indust Manage Rev 7:76–92, 1965, Industrial Manage Rev 7:83–88, 1966), and Frankel and Hawkins (J Finance 30:207–210, 1975) are representatives of the static stochastic models. While the dynamic stochastic models can be further divided into structural models and reduced-form models, respectively. For the structural models, Brennan and Schwartz (J Finance 32:1699–1716, 1977, J Financ Quantitat Analysis 15(4): 907–929, 1980) and Ingersoll (J Financ Econ 4:289–321, 1977a; J Finance 2: 463–478, 1977b) are the pioneers, in which the firm's market value is assumed to follow a geometric Brownian motion. The more recent literature develop reduced-form models using stochastic stock price instead of firm's market value for the underlying of the convertibles. These include the works by McConnell and Schwartz (1986), Goldman Sachs (Valuing convertible bonds as derivatives.

C.-F. Lee (✉)

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: cflee@business.rutgers.edu

L.-J. Kao

Department of Accounting Information, Takming University of Technology, Taipei, Taiwan

e-mail: ljkao@gs.takming.edu.tw

P.-C. Wu

Department of Finance and Banking, Kainan University, Taoyuan, Taiwan

e-mail: pcwu@mail.knu.edu.tw

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,

https://doi.org/10.1007/978-3-030-91231-4_68

1563

Technical report, Goldman Sachs. Quantitative Strategies Research Notes, 1994), Ho and Pfeffer (*Financ Analysts J* 52:35–44, 1996), Tsiveriotis and Fernandes (*J Fixed Income* 8(2):95–102, 1998), Davis and Lischka (Convertible bonds with market risk and credit default. Working paper, Tokyo Mitsubishi International Plc, 1999), and Hull (2000).

Keywords

Convertible debt valuation · Graphical method · Calculus model · Structural models · Reduced-form models · Geometric Brownian motion

JEL Classification

C52 · G12 · G31 · G35

Abbreviations

DL:	Deep Learning
ML:	Machine Learning
VaR:	Value-at-Risk
EVT:	Extreme Value Theory
LPA:	Local Parametric Approach
MLP:	Multilayer Perceptron
LSTM:	Long Short-Term Memory
GPD:	Generalised Pareto distribution
QMLE:	Quasi-Maximum Likelihood Estimation
AUC:	Area Under the roc Curve

68.1 Introduction

A convertible bond which is entitled to a set of fixed claims is a bond that can be converted into common stocks at the owner's option subject to conditions specified in the indenture. Thus, convertible bonds are two instruments in one: one is the straight debt instrument, which guarantees a fixed payment for the life of the bond. The second instrument is the equity portion, which allows the bondholder to reap the gains due to increase in the equity value by conversion. By issuing convertible bonds, the firms can lower its cost of debt funding compared to straight debt alone (Somanath 2011). For investors, convertible bonds offer greater stability of income than common stock.

There are two varieties of convertible bonds: callable versus non-callable convertibles. Usually, callable convertibles have the same provisions as their non-callable counterparts, except that they are subject to redemption by the issuing company at the prevailing call price prior to maturity. In most cases, the call provision is standard in indentures and is especially important in case of convertibles. As one can think of a non-callable convertible as a special case of callable convertible with sufficiently large call price, in this study, convertibles with call provision are always assumed.

In general, the price of a convertible is determined by three state variables and four decision variables, respectively (Choi et al. 2009). The three state variables are: (1) the price behavior of the firm's market value or common stock; (2) the value of the corresponding straight bond; (3) The credit risk of the issuing firm. The four decision variables are: (1) the call policy of a company with convertibles outstanding as when to announce a call, the call price for redemption payment, the specification of the call notice, . . . etc.; (2) the conversion policy includes the date of conversion, the conversion price as how much one share of common stock costs at conversion, or equivalently, the conversion ratio as how many shares of common stocks one debenture is convertible; (3) the put provision that allows the investor to sell the convertible bond to the issuer for specific prices on specific dates prior to maturity; (4) the capital structure of the firms, for example, whether there are any other senior claims of the firms.

In literature, various assumptions are employed regarding the aforementioned three state variables. The assumptions span from deterministic to stochastic. By assuming a constant interest rate r and deterministic stock price growing at a constant rate g , the graphical method of Brigham (1966) is a representative of the deterministic approach. To describe the dynamics of these state variables more rigorously, stochastic models are developed, which can be further divided into static and dynamic stochastic models, respectively. In the static stochastic models, one considers firm's market value (or stock price) on a specific date in the future, while in the dynamic stochastic models, one considers firm's market value (or stock price) for a time period in the future. The calculus models proposed by Baumol, Malkiel, and Quandt (1960), Poensgen (1965, 1966), and Frankel and Hawkins (1975) are representatives of the static stochastic models.

Both the deterministic and static stochastic approaches ignore the possibility of bankruptcy by the issuing firm so that the convertible bonds might be mispriced. To circumvent this problem, dynamic stochastic models are developed, and huge amount of works can be classified in this category, which can be divided into structural and nonstructural models, respectively, to analyze the valuation of convertibles along with optimal call, bankruptcy, and conversion strategies. The works by Brennan and Schwartz (1977) and Ingersoll (1977a, b) are the pioneers of one-factor structural models, in which the firm's market value following a geometric Brownian motion is considered. Later Brennan and Schwartz (1980) proposed a two-factor structural model that includes the stochastic interest rate. Nyborg (1996) extends Brennan and Schwartz's two-factor structural model by including a put provision and floating coupons feature to evaluate a convertible. In Liao and Huang (2006), a one-factor structural model with an endogenous strategic default is developed based on Leland (1994). The model is further characterized by incorporating two other essential elements of a convertible including the tax benefits due to coupon payments and bankruptcy costs that are closely connected with the bond issuer's capital structure.

The common feature of structural models is the assumption that default occurs when the firm's market value becomes too low to meet its default barrier. Nevertheless, due to the fact that the firm's market values are not observable, the structural models are subject to the criticism that parameter estimation is difficult (Ayache et al.

2003). In contrast to the structural models, the more recent literature use stochastic stock price instead of a firm's market value for the underlying of the convertibles, where credit risk is modeled in two alternative ways. One alternative is to adopt a riskier discount rate that is obtained by including an exogenously determined spread to the risk-free Treasury rate in the valuation of convertible (McConnell and Schwartz 1986; Ho and Pfeffer 1996; Tsiveriotis and Fernandes 1998). The other alternative is based on the reduced-form models for the pricing of risky debts (Jarrow and Turnbull 1995; Duffie and Singleton 1999; Madan and Unal 2000). These include the works by Goldman Sachs (1994), Davis and Lischka (1999), Takahashi et al. (2001), Ayache et al. (2003), and Lau and Kwok (2004). In these literatures, default events are considered occurring exogenously and the probability of a default event over the next short time interval is determined by its hazard rate. In Davis and Lischka (1999) and Takahashi et al. (2001), the hazard rate of default is considered as an additional state variable by an appropriate stochastic process.

This work reviews the literature on convertible debt valuation. Inadequacies in the historical and current valuation methods are highlighted. The outline of the chapter is as follows. In Sect. 68.2, the graphical model with deterministic assumptions is stated. Section 68.3 states the calculus models with static stochastic stock price and/or straight bond's price. Section 68.4 gives the structural models by modeling the dynamics of a firm's market value as a stochastic process. In Sect. 68.5, the models with dynamic stochastic stock price are stated. In Sect. 68.6, numerical approaches, for example, a tree approach, for the valuation of convertible bonds in reduced-form models is given. Section 68.7 concludes.

68.2 Deterministic Approach: Graphical Model

Brigham (1966) proposed a descriptive graphical model for the valuation of a callable convertible where the stock price is assumed to grow at a constant rate g and therefore the time- t conversion value satisfies

$$C_t = P_0(1 + g)^t R \quad (1)$$

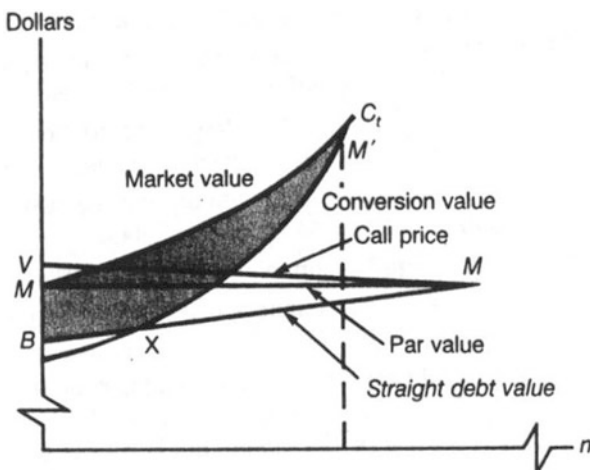
where R is the conversion ratio, P_0 is the initial stock price. By assuming a constant market interest rate i , the time- t straight-debt value is

$$B_t = \sum_{k=1}^{T-t} \frac{I}{(1+i)^k} + \frac{M}{(1+i)^{T-1}}$$

where M is the par value of the bond, T is the maturity date, and I is the coupon paid each period.

This model is presented in Fig. 1. Logically, the value of the convertible bond cannot be less than its straight-debt as well as conversion value. Therefore, the higher of the conversion value curve CC_t and straight-debt value curve BM forms a floor curve (BXC_t) for the market value of the convertible represented by the curve

Fig. 1 Hypothetical model of a convertible years' bond.
(Reprinted by permission from Brigham (1966))



(MM'). At relatively low conversion value, the convertible's floor curve is represented by the straight debt value, BX . Above the segment BX , the shaded area refers to the value attached to the conversion feature of the security. Thus, at the lower levels of equity prices, the convertible's market value is a composite of its straight debt value and a premium for the conversion feature. Moving to higher levels of equity prices, there exists a premium between the conversion value XC_t and the convertible's market value MM' due to the risk exposure reduction by holding the convertibles. As the conversion value increases to exceed the call price VM , the convertible's market value approaches M' where the convertible's market value intersects with the conversion value. The reasons for the intersection are threefold: (i) the likelihood of a call increases as the conversion value increases; (ii) loss protection against interest rate risk diminishes; (iii) the dividend yield of the common stock exceeds the fixed coupon yield of the convertible as the conversion value increases. In summary, a convertible's market price is a complex interaction between its straight debt value and a premium associated with the expected future market price of the underlying equity security. This complex interaction, however, is not explicitly specified in Brigham's work. In the following, the interaction is accounted for by random characterization of the stock price and/or the straight debt value.

68.3 Static Stochastic Stock and/or Straight Bond Prices

Baumol, Malkiel, and Quandt (1960) present a convertible-security valuation model that separates the interaction between debt and equity attributes in the valuation of convertible securities by considering the time- t value $C = \max(C_s, C_b)$ of a convertible as

$$C_s = ps + \int_0^{\bar{B}/ps} f(i, t_0) [\bar{B} - i(t)ps] di(t), \quad (2)$$

$$C_b = \bar{B} + \int_{\bar{B}/ps}^{\infty} f(i, t_0) [i(t)ps - \bar{B}] di(t), \quad (3)$$

where $\bar{B}$ = equivalent straight debt value; s = number of equity securities issued upon conversion; p = price of the equity security at present t_0 ; $i(t)$ = date- t to date- t_0 equity security's price relative; $f(i, t_0)$ = subjective probability assessment of $i(t)$ at t_0 , when $t > t_0$. Equation (2) specifies that C_s is composed of two values: the conversion value ps and the insurance value of the debt character as the stochastic price relative $i(t)$ varies between 0 and $\bar{B}/ps$ at the end of time horizon t . Equation (3) states that C_b is a combination of two values: the straight debt value $\bar{B}$ and the value assigned to the conversion feature as the stochastic price relative $i(t)$ varies between $\bar{B}/ps$ and ∞ at the end of time horizon t .

Later, Baumol, Malkiel, and Quandt (1960) extended their previous model by introducing the ex ante notion that the convertible security holder is interested in the future conversion value of the convertible security as follows:

$$C = \int_0^{\infty} i(t)psf(i, t_0)di(t) + \int_0^{\bar{B}/ps} f(i, t_0) [\bar{B} - i(t)ps] di(t), \quad (4)$$

In this model, the first term represents the conversion value in all possible states of the price relative $i(t)$ from 0 to ∞ , while the second term represents the insurance value of the debt character as the stochastic price relative $i(t)$ varies between 0 and $\bar{B}/ps$.

Both in Brigham (1966) and Baumol, Malkiel, and Quandt (1960), the market interest rate is flat and bond's straight debt value was treated as an exogenously determined constant. By including the nature of interest-rate changes, an explicit characterization of the random nature of a convertible security's bond value would lead to a robust model of the market setting of convertible securities. In Poensgen (1965), both the stochastic stock price and stochastic straight debt value are explicitly considered. Be more specific, on an arbitrarily date when conversion is possible, Poensgen (1965) suggests the expected price of the convertible security as

$$E(P) = \int_0^{\infty} \left[y \int_0^y h\left(\frac{x}{y}\right) dx + \int_y^{\infty} xh\left(\frac{x}{y}\right) dx \right] g(y) dy, \quad (5)$$

where x = the stock price (or conversion value); y = the straight debt value; $h(x/y)$ = the conditional probability of x for a particular bond value y ; $g(y)$ = the probability of occurrence of the bond value y (unconditional). Note that the probability of joint occurrence of the stock price x and bond value y is $h(x, y) = h(x/y)g(y)$. It can be shown that Eq. (5) is logically equivalent to the following:

$$E(P) = \underbrace{\int_0^\infty \left[\int_0^\infty xh(x,y)dx \right]}_{\text{Expected stock value}} + \underbrace{\int_0^y (y-x)h(x,y)dx}_{\text{Value of floor guarantee}} dy, \quad (6)$$

$$E(P) = \underbrace{\int_0^\infty yg(y)dy}_{\text{Expected straight debt value}} + \underbrace{\int_0^\infty \int_y^\infty (x-y)h(x,y)dx dy}_{\text{Expected value of the conversion option}}, \quad (7)$$

Equation (6) specifies that the value of the convertible security is the expected stock price plus the value attached to the floor guarantee for a stochastic straight debt value, y . This formulation is identical to Eq. (4) by Baumol, Malkiel, and Quandt's (1960) model except for the assumption by Poensgen (1965) that y is a random variable. Alternatively, in Eq. (7), the first term evaluates the value of a straight debt, while the second term evaluates the value of a warrant with exercise price y by integrating over all possible straight-debt value, y .

By showing that there was no statistically significant correlation between bond yield variability and stock prices, Poensgen (1965) claims that equation Eq. (7) is analytically equal to

$$E(P) = \int_0^\infty yg(y)dy + \int_y^\infty (x-y)f(x)dx, \quad (8)$$

where $f(x)$ is the probability density function of x . Based on Eq. (8), Poensgen (1965) suggests that the distribution of stock prices, x , is truncated on the left by the straight-debt value. This assertion facilitates Poensgen's (1965) subsequent empirical work dealing with the truncated distribution of convertible security returns. Nevertheless, the integration in Eq. (8) does not equal to that in Eq. (7). Therefore, the subsequent work of Poensgen (1965) on the truncated distribution of convertible security returns is not of great help.

Later, Frankel and Hawkins (1975) adopted Poensgen's valuation model for convertible securities to explore the link between the beta of a convertible security and the beta of the underlying common stock by assuming independence of the stock and straight bond prices so as to obtain

$$E(CB) = \int_0^y yh(x)dx + \int_y^\infty xh(x)dx, \quad (9)$$

where $E(CB)$ = expected price of the convertible security; y = straight-debt value, x = log of the stock price relative; $h(x)$ = log-normal distribution of stock price relatives. By assuming a constant straight debt value and following the previous

direction of Poensgen (1965) by asserting that the probability density function of the log of the stock price relatives are truncated at a constant value a , Frankel and Hawkins (1975) presents the expected log return on the convertible bond as:

$$E_{en(i(t))CB} = a \int_{-\infty}^a \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(1/2)[(x-\mu_x)/\sigma_x]^2} dx + \int_a^{\infty} x \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(1/2)[(x-\mu_x)/\sigma_x]^2} dx. \quad (10)$$

Frankel and Hawkins (1975) proceed to derive an explicit form for the beta of a convertible security based on the beta of the underlying common stock. This relationship is presented as follows:

$$\text{Beta} = \rho \frac{\sigma_x}{\sigma_m} \left[1 - \phi \left(\frac{a}{\sigma_x} \right) \right], \quad (11)$$

where ρ = correlation coefficient between x and the market m ; σ_x = standard deviation of common stock return; σ_m = standard deviation of market return; ϕ = the cumulative normal function. It is noted by Frankel and Hawkins that (11) implies that the beta of a convertible security is less than or equal to that of its underlying common stock. Thus, Frankel and Hawkins (1975) derive an upper bound for the convertible security's beta. Nevertheless, the result by Frankel and Hawkins (1975) in Eq. (11) is false as Eq. (9) is falsely derived. Further efforts need to be done for a valid relationship between the beta of a convertible security and that of its underlying common stock.

The next major advance in convertible-security valuation evolved from the option pricing theory originated by Black and Scholes (1973) and the risky zero-coupon bond's pricing by Merton (1974). The following section considers the extension of structural models for the pricing of convertible bonds.

68.4 Structural Models: Dynamic Stochastic Firm's Market Value

In the structural models, default is either an exogenously determined event when the firm's market value falls below the principle of its debts (Merton 1974; Black and Cox 1976; Longstaff and Schwartz 1995; Briys and Varenne 1997), or an event when the firm's market value falls below an endogenously determined barrier due to the strategic decision of the bond's issuer (Leland 1994; Leland and Toft 1996). In either type of the aforementioned structural models, the determinant of a default event is the firm's market value: default occurs if the market value falls below to the default barrier for the first time before the maturity or at maturity.

Due to the aforementioned feature, one advantage of structural models is that it captures the dilution effect as convertible bonds are converted into common stocks and then be expressed as a fraction of the firm's market value (Schönbuchner 2003). With this advantage, Ingersoll (Ingersoll Jr 1977a) is able to derive the optimal conversion strategy for the investor and call strategy for the firm with capital structure consisting of equities and convertible bonds only in a perfect, frictionless market with no dividends and constant conversion terms. According to Ingersoll (Ingersoll Jr 1977a), the optimal conversion strategy satisfies the following lemma:

Lemma 1 If the perfect market, no dividends, and constant conversion terms assumptions are valid, then a callable convertible will never be converted except at maturity or a call.

With the additional assumptions of no call notice and flat interest rate, Ingersoll (Ingersoll Jr 1977a) gave the optimal call strategy for a company as

Lemma 2 If the conditions for the above lemma hold, the optimal call strategy is to call when the firm's market value reaches the call price divided by the dilution factor γ , the fraction of the equity that the bond holders possess if they convert.

Using the above Lemmas and by assuming the firm's market value following a geometric Brownian motion, Ingersoll (1977a) first decomposed the price of a non-callable convertible bond G into a straight discount bond F (with the same principal as the convertible bond) and a warrant W as

$$G(V, \tau; B, 0, \gamma) = F(V, \tau; B, 0) + W(\gamma V, \tau; B) \quad (12)$$

where V is the firm's market value, B is the balloon payment, τ is the time to maturity, c is coupon payment, γ is the dilution factor, $W(\gamma V, \tau; B)$ is the warrant price with underlying price γV and exercise price B . The initial condition for W satisfies

$$W(\gamma V, 0; B) = \max(\gamma V_T - B, 0)$$

where V_T is the value of firm at maturity. Ingersoll (1977a) is able to derive analytically the values of the straight discount bond F and the warrant W using the option pricing theory originated by Black and Scholes (1973) and extended by Merton (1974) for a risky debt. In the case of a callable convertible bond, the price H is decomposed into a straight discount bond F , a warrant W and an additional term representing the cost of the call which reduces the value of the callable convertible bond relative to its non-callable counterpart as

$$\begin{aligned} H(V, \tau; K(\tau), 0, \gamma) &= F(V, \tau; B, 0) + W(\gamma V, \tau; B) \\ &\quad + Z^{2(r-p)/\sigma^2} [F(\gamma V', \tau; B', 0) - F(\gamma V', \tau; B'/\gamma, 0)] \end{aligned} \quad (13)$$

where $K(\tau)$ is the call price of the convertible and $Z = K(\tau)/\gamma V$, $B' = Be^{(r-\rho)\tau}$, $V' = Ve^{(r-\rho)\tau}$. In another paper by Brennan and Schwartz (1977), the partial differential equation for the price of a convertible with call provisions and dividends is derived based on the following conversion and call strategies.

Lemma 3 It will never be optimal to convert a convertible bond except immediately prior either to a dividend date or to an adverse change in the conversion terms, or maturity.

Lemma 4 The firm's optimal call strategy is to call the bond as soon as its value, if it is not called, is equal to the call price.

Later, in addition to the simplistic capital structure assumed by Ingersoll (Ingersoll Jr 1977a) and Brennan and Schwartz (1977) consisting solely of equity and convertible bonds, Brennan and Schwartz (1980) include senior straight debts in a firm's capital structure and develop a two-factor structure model with stochastic firm's market value and interest rate for the valuation of a convertible bond. Brennan and Schwartz (1980) found that often the additional factor representing stochastic interest rates had little impact on the prices of convertible bonds. Later, Nyborg (1996) extends the two-factor model of Brennan and Schwartz (1980) with a put provision and floating coupons feature.

In contrast to the aforementioned structure models in which the default barrier is exogenously determined by limited liability, that is, default occurs when the firm's market value falls below the principal value of its total outstanding debt, Liao and Huang (2006) develop a structural model whose default barrier is endogenously determined due to the strategic decision of the bond's issuer (Leland 1994). In this setting, Liao and Huang's model (2006) derive the optimal call, conversion, and bankruptcy strategies by maximizing the equity value that includes the refunding cost when the convertible is called and the bankruptcy cost when the firm is bankrupt, respectively.

Liao and Huang (2006) took two steps to derive the price of a callable convertible when a firm has a capital structure with a callable convertible, a perpetual straight bond of equal priority, and common stocks. Specifically, let the par value of the convertible that continuously pays constant coupon flows, C_1 , with infinite maturity be P_1 ; and the par value of the perpetual straight bond that pays continuous coupon payments C_2 be P_2 . In the first step of Liao and Huang (2006), the prices of a non-callable convertible and of a call-forcing convertible that does not allow voluntarily conversion by the bondholders are given, respectively. The price of a non-callable convertible is given by

$$\begin{aligned} \text{NCCB}(V; V_{B1}, V_{con}) = & \frac{C_1}{r} + \left(\frac{(1-\alpha)P_1}{P_1 + P_2} V_{B1} - \frac{C_1}{r} \right) G_{\tau B1} \\ & + \left(\gamma V_{con} - \frac{C_1}{r} \right) G_{\tau con} \end{aligned} \quad (14)$$

where $G_{\tau_{B1}}$ is the state price of bankruptcy, and $G_{\tau_{con}}$ is the state price of conversion occurring prior to bankruptcy, V_{B1} represents the bankruptcy trigger of the issuer, and V_{con} denotes the voluntary conversion trigger of the bondholders. The price of a call-forcing convertible is given by

$$\begin{aligned} \text{CFCB}(V; V_{B2}, V_{call}) = & \frac{C_1}{r} + \left(\frac{(1-\alpha)P_1}{P_1 + P_2} V_{B2} - \frac{C_1}{r} \right) H_{\tau_{B2}} \\ & + \left(\max(\gamma V_{call}, K) - \frac{C_1}{r} \right) H_{\tau_{call}} \end{aligned} \quad (15)$$

where $H_{\tau_{B2}}$ is the state price of bankruptcy before the call, and $H_{\tau_{call}}$ is the state price of call occurring prior to bankruptcy, K is the call price, V_{B2} represents the bankruptcy trigger of the issuer, and V_{call} denotes the call trigger of the issuer.

Second, by viewing the tax benefits due to coupon payments and bankruptcy costs as two contingent claims upon the firm's market value, the optimal bankruptcy trigger $V_{B1}^*(V_{B2}^*)$, the call trigger V_{call}^* , and conversion trigger V_{con}^* are derived. Now the price of a callable convertible can be obtained readily either from that of the non-callable convertible in (14) or from that of the call-forcing convertible in (15), which are otherwise identical, according to the following two lemmas.

Lemma 5 Whenever it is optimal to voluntarily convert a non-callable convertible, it will also be optimal to convert a callable convertible that is otherwise identical.

Lemma 6 Whenever it is optimal to call back a call-forcing convertible, it will also be optimal to call back a callable convertible that is otherwise identical.

68.5 Dynamic Stochastic Stock Price: Spread and Reduced-Form Approaches

The use of firm's market value as the underlying for the convertible bond, though self-consistent in that all the securities and therefore the firm's capital structure are addressed, can be impractical due to the fact that the firm's market value is unobservable and the number of parameters used is large. Instead, practitioners often use the equity price as the underlying for the valuation of convertible bonds.

The approaches using equity prices as the underlying can be classified into two categories. In the first category, a riskier discount rate obtained by adding an exogenously determined spread to the risk-free Treasury rate is used in the valuation of a convertible (McConnell and Schwartz 1986; Ho and Pfeffer 1996; Tsiveriotis and Fernandes 1998; Jarrow and Protter 2004). We call this type of models as spread approach. In McConnell and Schwartz (1986), a one-factor model in which the stock price follows a lognormal process with constant dividend yield or constant dividend growth rate is developed. To account for the default risk, however, McConnell and

Schwartz (1986) adopt a risky discount rate in an arbitrary fashion. Ho and Pfeffer (1996) developed a two-factor model that includes a stochastic interest rate. The credit risk of the convertible is captured by the option-adjusted spread (OAS) that is estimated based upon the prices of bonds of the same sector and rating groups. Noting that different components of a convertible might have different credit risks, Tsiveriotis and Fernandes (1998) used different credit spreads for the valuation of a convertible bond by dividing a convertible bond into a cash account and an equity account. The cash account entitled to the bondholders all the cash flows receives; while the equity account entitled to the bondholders the payments in equity. The cash account is discounted at the risky rate as defined above, while the equity account is discounted at the risk-free rate. In Goldman Sachs (1994), two extreme situations are considered for the credit spreads. The first situation is when the stock price is far above the conversion price and conversion is certain to be exercised, while the second situation is when the stock price is far below the conversion price so the value of a convertible is close to a risky corporate straight bond. For the first situation, the risk-free rate is used; while a risky rate which includes an exogenously determined spread, that is, the issuer's credit spread, is used for the second situation.

The aforementioned models suffer from the same drawback that the possibility of a sudden loss event that drives the firm to default is not included in the model. Furthermore, they do not explicitly specify the terms, for example, the recovery, in the event of a default by the issuing firm. The models in the second category that are based on the reduced-form models for the pricing of risky debts emerged to account for the possibility of a sudden loss event (Jarrow and Turnbull 1995; Duffie and Singleton 1999; Madan and Unal 2000).

In the reduced-form model setting, default is exogenously driven with probability over the next short time period determined its hazard rate. In Ayache, Forsyth, and Vetzal (2003) and Lau and Kwok (2004), one-factor models assuming a Poisson default process with a constant hazard rate are developed. In case of a default, a sudden stock price jump and various assumptions concerning recovery on default are considered in these models. In Lau and Kwok (2004), the impacts of the hard call constraint that restrains a call during the early life of the bond, of the soft call constraint that requires the stock price to be above certain trigger price to initiate a call, and of a call notice period are examined. They attributed the "delayed call phenomena" to the underestimation of the critical call price due to inaccurate modeling of the contractual provisions.

More advanced reduced-form models consider the default hazard rate as an underlying of the convertible bond to account for the potential default risk. In Davis and Lischka (1999), a three-factor model is specified for the dynamics of equity price, the interest rate, and the default hazard rate. The default hazard rate is included in the drift term of the equity price process to capture the possibility that the equity price might jump to zero suddenly in the event of default. Three possible models are considered in Davis and Lischka (1999): (1) A stochastic equity price process, a mean-reverting interest rate process, and a deterministic hazard rate; (2) a stochastic equity price process, a mean-reverting hazard rate process, and a deterministic interest rate; and (3) a stochastic equity price

process, a mean-reverting interest rate process, and a hazard rate following a perfectly negatively correlated arithmetic Brownian motion with respect to the equity process.

In another paper by Takahashi et al. (2001), a reduced-form model for convertible bonds is constructed based on Duffie and Singleton (Duffie and Singleton 1999). In Takahashi et al. (2001), the risky discount rate $R(t)$ is determined by the risk-free deterministic interest rate $r(t)$, the default hazard rate $\lambda(S, t)$, and the fractional loss rate $L(t)$ at default in the form

$$R(t) = r(t) + L(t)\lambda(S, t) \quad (16)$$

Here the default hazard rate $\lambda(S, t)$ is assumed to be a decreasing function of the stock price S . By allowing conversion before maturity, the value of a convertible is formulated as an optimal stopping time problem as

$$\sup_{\tau \in \Gamma} E_t^Q \left[e^{-\int_t^\tau R(u) du} 1_{(\tau < T)} \alpha S_T + e^{-\int_t^T R(u) du} 1_{(\tau = T)} \max(\alpha S_T, F) + \sum_{t < T_i < \tau} c_i e^{-\int_t^{T_i} R(u) du} \right] \quad (17)$$

where Γ denotes the feasible set of conversion times, F is the face value, c_i is the coupon payment at time T_i , and α is the conversion ratio.

68.6 Numerical Methods

Due to the complexity of the convertible's contractual provisions, there might be no analytical solutions for the valuation of the security, and some numerical approaches need to be employed. In this section, prevalent numerical approaches that are often used in the valuation of convertible bonds are stated.

68.6.1 Solving Partial Differential Equations

If the firm's market value V follows a lognormal Ito process in a one-factor structural model, according to Merton (1974), then the value f of the firm's convertible satisfies the partial differential equation

$$\frac{1}{2} \sigma^2 V^2 f_{VV} + (rV - C) f_V - rf - f_\tau + c = 0 \quad (18)$$

where C is the cash payments to all claims of the firm, c is that portion of payments to the claim under consideration, τ is the time to maturity. If further incorporating a second factor, that is, the stochastic interest rate r , that follows a diffusion process

$$dr = \theta(\mu_r - r)dt + r\sigma_r dW$$

then the value of the convertible satisfies

$$\begin{aligned} \frac{1}{2}\sigma_V^2 V^2 f_{VV} + \frac{1}{2}\sigma_r^2 r^2 f_{rr} + (rV - C)f_V + [\theta(\mu_r - r) - r\lambda\sigma_r]f_r \\ + \rho V r \sigma_r \sigma_V f_{Vr} - rf - f_\tau + c = 0 \end{aligned} \quad (19)$$

where λ is the market price of interest rate risk. By specifying boundary conditions determined by the convertible's contractual provisions, Eqs. (18) and (19) can be solved using relevant numerical methods. For example, in Brennan and Schwartz (1977, 1980), the finite difference method is used to solve (18) and (19) to obtain the value of a convertible security.

Instead of using a firm's market value, if considering a one-factor model in which the firm's stock price S following a lognormal Ito process of the form

$$dS = \mu(S, t)Sdt + \sigma SdW$$

where $\mu(S, t)$ is the drift term, σ is the volatility, and W is a standard Brownian motion. Then the value f of a convertible is governed by a Black-Scholes formula of the form

$$\frac{1}{2}\sigma^2 S^2 f_{SS} + \mu S f_S - Rf - f_\tau + c = 0 \quad (20)$$

where μ is the drift term, R is the risky discount rate, c is the cash payments to the convertible, and τ is the time to maturity.

In Tsiveriotis and Fernandes (1998), the drift term $\mu(S, t)$ is the growth rate of the stock and the risky discount rate R is the sum of the risk-free interest rate and credit spread. By dividing a convertible into a cash and equity account, in which boundaries associated with the call, put and conversion options for the equity account are also the boundary conditions for the cash account, Tsiveriotis and Fernandes (1998) write down two coupled partial differential equations in the form of (20) for the cash and equity accounts respectively and finite difference approach is used to solve the coupled PDEs.

In Takahashi et al. (2001), the drift term $\mu(S, t) = r(t) - q + \lambda(S, t)$, where $r(t)$ is the risk-free interest rate, q is the dividend rate, and $\lambda(S, t)$ is the default hazard rate assumed to be a decreasing function of the stock price S , while the discount rate R is given in (16). Takahashi et al. (2001) solved the partial differential Eq. (20) using finite difference method to obtain a solution for (20). In Ayache, Forsyth, and Vetzal (2003), the discount rate R is $r(t) + \lambda$; while the drift term $\mu(S, t) = r(t) - q + \lambda\eta$, where $r(t)$ is the risk-free interest rate, λ is the constant default hazard rate, and $S^-, (1-\eta)S^-$, $0 \leq \eta \leq 1$, denote the stock prices before and after a jump event, respectively. Ayache, Forsyth, and Vetzal (2003) present the PDEs in terms of a system of linear complementarity problem and a partially implicit method is used to decouple the system of linear complementarity problems at each time step. In Lau and Kwok

(2004), the drift term $\mu(S, t)$ and the discount rate R are similar to those of Takahashi et al. (2001). Lau and Kwok (2004) apply a dynamic programming procedure at each node of an effective explicit finite difference algorithm with consideration of the interaction of the call and conversion features.

68.6.1.1 The Tree Approach

The most widely used numerical approach for the pricing of a convertible is the tree approach. The approach has the capability to handle various convertible bond provisions. The main procedure of the tree approach for convertible bond pricing is stated as follows:

First, a binomial tree of the stock price in a risk neutral world is built according to the Cox, Ross, and Rubinstein (1979). The tree expanding method is already the standard approach to pricing an option. Each node on the tree represents a possible future stock price. Then a backward induction procedure starting from the maturity is applied, a maximum convertible bond value on each node is decided according to the stock price and the action taken according to the provisions, for example, convert, put, call, hold, or redeem the bond. Finally the current value of convertible bond can be calculated. The credit risk of the convertible bond is considered by setting a risky discount rate with the credit spread of the issuer (Goldman Sachs 1994) or adding a default node (Hull 2009) with the default probability which decided by the default intensity of the issuer.

More complex model with multi-factors, such as adding a stochastic interest rate, or with time-varying default intensity, volatility, and interest rate, is also applicable with tree model (Ho and Pfeffer 1996; Hung and Wang 2002; Ammann et al. 2003, 2008; Carayannopoulos and Kalimipalli 2003; Chambers and Lu 2007). Ho and Pfeffer (1996) employed a two-dimensional binomial tree for the stochastic stock price and interest rate in their two-factor model, in which all the cash flows are discounted at the risky rate (i.e., risk free plus credit spread). Correlated default intensity and stock price is discussed in Andersen and Buffum (2003). However, Bernnan and Schwartz (1980) and Goldman Sachs (1994) indicated that the major factor which decides the convertible bond value is the equity price. Adding additional factor, such as dynamic interest rate, into consideration had little impact on the value of convertible bond.

68.7 Conclusion

This chapter reviewed the various models developed for the valuation of convertible bonds in which the call, voluntary conversion, and bankruptcy can occur at any time of the bond's duration: on the one hand, the shareholders determine the optimal call and bankruptcy policies so as to maximize the equity value, but on the other hand, the convertible bondholders determine the optimal voluntary conversion strategy so as to maximize the convertible bond's value. This gives rise to a dual option in that the bondholders' optimal voluntary conversion strategy depends on the firm's call strategy, and the firm's optimal call strategy in turn depends on the bondholders'

voluntary conversion strategy. The valuation problem would be considerably complex without any simplification assumptions.

Study further takes account of the bankruptcy costs, tax benefits, and capital structure of the bond issuer as well as the refunding costs and a call notice period can be found in Liao and Huang (2005). In Lee et al. (2009), the association between investor protection and the security design of convertibles, as measured by the expected probability of converting the convertibles at maturity, is examined. For convertible bonds with nonstandard features, such as the refix clause that alters the conversion ratio (shares per bond) or conversion price, subject to the share price level on certain days between issue and expiry, the convertible bond's coupons can be allowed to change with time, or at conversion the bondholder receives a combination of shares and cash instead of just shares, more advanced models need to be developed to take into account of the versatile designs of the convertible bonds. For the summary, see Grimwood and Hodges (2002), Loncarski, Horst, and Veld (2009). An empirical comparison of the various valuation methods of convertible bonds, see Zabolotnyuk, Jones, and Veld, (2010).

Acknowledgments This work had been finished with suggestions and comments from Professor Lee and his colleague.

References

- Ammann, M., A. Kind, and C. Wilde. 2003. Are convertible bonds underpriced? An analysis of the French market. *Journal of Banking and Finance* 27: 635–653.
- . 2008. Simulation-based pricing of convertible bonds. *Journal of Empirical Finance* 15: 310–331.
- Andersen, L., and D. Buffum. 2003. Calibration and implementation of convertible bond models. *Journal of Computational Finance* 7 (1): 1–34.
- Ayache, E., P.A. Forsyth, and K.R. Vetzal. 2003. The valuation of convertible bonds with credit risk. *Journal of Derivatives* 11: 9–29.
- Baumol, W.J., B.G. Malkiel, and R.E. Quandt. 1960. The valuation of convertible securities. *The Quarterly Journal of Economics* 80: 48–60.
- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–654.
- Brennan, M.J., and E.S. Schwartz. 1977. Convertible bonds: Valuation and optimal strategies for call and conversion. *The Journal of Finance* 32: 1699–1716.
- Brennan, M.J., and E. Schwartz. 1980. Analyzing convertible bonds. *Journal of Financial and Quantitative Analysis* 15 (4): 907–929.
- Brigham, E.F. 1966. An analysis of convertible debentures: Theory and some empirical evidence. *Journal of Finance* 21: 35–54.
- Briys, E., and F. de Varenne. 1997. Duration of Risky Debts: Integrating Early Default and Absolute Priority Rule Deviations. Working Paper, Federation Francaise des Societes d'Assurances.
- Carayannopoulos, P., and M. Kalimipalli. 2003. Convertible bond prices and inherent biases. *The Journal of Fixed Income* 13: 64–73.

- Chambers, D.R., and Q. Lu. 2007. A tree model for pricing convertible bonds with equity, interest rate, and default risk. *Journal of Derivatives* 14: 25–46.
- Choi, D., M. Getmansky, and H. Tookes. 2009. Convertible bond arbitrage, liquidity externalities, and stock prices. *Journal of Financial Economics* 91: 227–225.
- Cox, J.C., S.A. Ross, and M. Rubinstein. 1979. Option pricing: A simplified approach. *Journal of Financial Economics* 7: 229–263.
- Davis, M., and F. Lischka. 1999. Convertible bonds with market risk and credit default. Working paper, Tokyo Mitsubishi International Plc. Tokyo, Japan.
- Duffie, D., and K. Singleton. 1999. Modeling term structure of defaultable bond. *Review of Financial Studies* 12: 687–720.
- Frankel, A.W., and C.A. Hawkins. 1975. Beta coefficients for convertible bonds. *Journal of Finance* 30: 207–210.
- Grimwood, R., and S. Hodges. 2002. *The valuation of convertible bonds: A study of alternative pricing models*, Working paper. Warwick University. Coventry CV4 7AL, United Kingdom
- Ho, T., and D. Pfeffer. 1996. Convertible bonds: Model, value attribution and analytics. *Financial Analysts Journal* 52: 35–44.
- Ho, T., and D. Pfeffer. 1996. Convertible bonds: Model, value attribution and analytics. *Financial Analysts Journal* 52: 35–44.
- Hull, J.C. and A. White. 2000. Valuing Credit Default Swaps II: Modeling Default Correlations, Working Paper in progress, University of Toronto.
- Hull, J.C. 2009. *Options, futures, and other derivatives*. 7th ed. Harlow: Pearson Education.
- Hung, M.W., and J.Y. Wang. 2002. Pricing convertible bonds subject to default risk. *The Journal of Derivatives* 10: 75–87.
- Ingersoll, J.E., Jr. 1977a. A contingent-claims valuation of convertible securities. *Journal of Financial Economics* 4: 289–321.
- . 1977b. An examination of corporate call policies on convertible securities. *Journal of Finance* 2: 463–478.
- Jarrow, R.A., and S. Turnbull. 1995. Pricing Options on Derivative Securities Subject to Credit Risk, *Journal of Finance* 50(1): 53–85.
- Jarrow, R.A., and P. Protter. 2004. Structural versus reduced form models: A new information based perspective. *Journal of Investment Management* 2: 34–43.
- Lau, K.W., and Y.K. Kwok. 2004. Anatomy of option features in convertible bonds. *Journal of Futures Markets* 24: 513–532.
- Lee, C.F., K.W. Lee, and G. Yeo. 2009. Investor protection and convertible debt design. *Journal of Banking & Finance* 33: 985–995.
- Leland, H.E. 1994. Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance* 49: 1213–1252.
- Leland, H.E., and K.B. Toft. 1996. Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *Journal of Finance* 51: 987–1019.
- Liao, S.L., and H.H. Huang. 2006. Valuation and optimal strategies of convertible bonds. *Journal of Futures Markets* 25: 895–922.
- Loncarski, I., J. Horst, and C. Veld. 2009. The rise and demise of the convertible arbitrage strategy. *Financial Analysts Journal* 26: 35–50.
- Longstaff and Schwartz. 1995. A simple approach to valuing risky fixed and floating rate debt, *Journal of Finance*, 50: 789–819.
- Madan, D. and H. Unal. 2000. A two-factor hazard rate model for Pricing Risky Debt and the Term Structure of Credit Spreads, *Journal of Financial and Quantitative Analysis*.
- Merton, R. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29: 449–470.
- Nyborg, K. 1996. The use and pricing of convertible bonds. *Applied Mathematical Finance* 3: 167–190.
- Poensgen, O.H. 1965. The valuation of convertible bonds, Part 1. *Industrial Management Review* 7: 76–92.

- . 1966. The valuation of convertible bonds, Part 2. *Industrial Management Review* 7: 83–88.
- Goldman Sachs. 1994. Valuing convertible bonds as derivatives. Technical report, Goldman Sachs. *Quantitative Strategies Research Notes*.
- Somanath, V.S. 2011. *International financial management*. I.K. International Publishing House Pvt. Ltd. New Delhi, India.
- Takahashi, A., T. Kobayashi, and N. Nakagawa. 2001. Pricing convertible bonds with default risk: A Duffie-Singleton approach. *The Journal of Fixed Income* 11: 9–20.
- Tsiveriotis, K., and C. Fernandes. 1998. Valuing convertible bonds with credit risk. *The Journal of Fixed Income* 8 (2): 95–102.
- Zabolotnyuk, Y., R. Jones, and C. Veld. 2010. An empirical comparison of convertible bond valuation models. *Financial Management* 39 (2): 675–705.

A Rationale for Hiring Irrationally Overconfident Managers

69

Oded Palmon and Itzhak Venezia

Abstract

The viability of managerial overconfidence is perplexing since it has been shown to lead managers to erroneous and costly decisions. This chapter addresses this issue by exploring the impact of managerial overconfidence on managerial effort, executive compensation, and the welfare of stockholders and managers. Overconfidence affects managerial effort directly and indirectly. The direct effect is that the optimal effort chosen by managers is positively related to their level of overconfidence. The indirect impact is through the influence on stockholders' choices of contract parameters. Thus, managerial overconfidence helps mitigate the well-known conflict of interest between managers and stockholders that induces managers to exert effort levels that are lower than the socially optimal levels. We construct a measure of the combined welfare of managers and stockholders and show that it is positively related to managerial overconfidence, thus providing an explanation to the persistence of this bias.

We thank Darius Palia, Orly Sade, and seminar participants at Rutgers University and Bocconi University for comments and suggestions. The financial support of The Sanger Family Chair for Banking and Risk Management, The Galanter Fund, The Mordecai Zagagi Fund, the Whitcomb Center for Research in Financial Services, and The School of Accounting, The Hebrew University are gratefully acknowledged.

O. Palmon (✉)

Rutgers Business, Rutgers University, Piscataway, NJ, USA

Rutgers Business, Rutgers University, Newark and New Brunswick, NJ, USA

e-mail: palmon@rbs.rutgers.edu

I. Venezia

School of Business, The Hebrew University, Jerusalem, Israel

The Hebrew University of Jerusalem, Jerusalem, Israel

The Academic College of Tel Aviv-Yaffo, Tel Aviv, Israel

e-mail: msvenez@mscc.huji.ac.il

Keywords

Behavioral biases · Incentive options · Managerial effort · Overconfidence

JEL CodesG3 · G34

69.1 Introduction

Confidence is a highly regarded personality trait in our society. Confident people are often greatly admired and lack of confidence is considered as an unattractive quality. Overconfidence is a well documented psychological bias whose consequences can vary.¹ Whereas overconfident individuals can bring about bold and imaginative innovations, their overconfidence may also lead them to make erroneous and costly decisions.

There are two main schools of thought regarding the definition and measurement of overconfidence. First, Alpert and Raiffa (1982), and their followers, concentrate on the calibration (or precision) of probabilities. They consider someone to be overconfident if her view of the precision of her estimate of some variable is too high, or put differently if she attaches too low a probability to the event that she may be wrong. This type of bias is akin to underestimating the standard deviation of some variables.² According to the other school of thought, people are considered overconfident if they are too optimistic regarding their accomplishments and attach too high a probability of success to their endeavors; that is, they overestimate the expected values of the results of their effort. The difference between the two schools of thought can be summed up in that the first one assumes that overconfident people overestimate their ability to predict the future whereas the second assumes that they overestimate their ability to change the future.

There is evidence that people overestimate their ability to do well on tasks (Frank 1935), are unrealistically optimistic about future events, expect good things to happen to them more often than to their peers (Weinstein 1980; Weinstein and Klein 2002; Kunda 1987), and are even unrealistically optimistic about pure chance events (Marks 1951; Irwin 1953; Langer and Roth 1975). Most individuals see themselves as better than the

¹See, e.g., DeBondt and Thaler (1995), Shefrin (2000), Barberis and Thaler (2003), Baker et al. (2006); and Camerer and Malmendier (2007).

²Lichtenstein et al. (1982), provide a review of the findings on overconfidence in many professions. For overconfidence among Physicians and nurses see: Christensen-Szalanski and Bushyhead (1981), and Baumann et al. (1991); among investment bankers: Staël von Holstein and Carl-Axel (1972); among engineers: Kidd 1970; among entrepreneurs: Cooper et al. (1988); among lawyers: Wagenaar and Keren (1986); among negotiators: Neale and Bazerman (1990); among accountants: Bar-Yosef and Venezia (2008); and among managers: Russo and Schoemaker (1992).

average person and most individuals see themselves better than others see them (Taylor and Brown 1988). They also rate their abilities and their prospects higher than those of their peers.³ For example, new business owners think that their business is more likely to succeed than other businesses like theirs' (Cooper et al. 1988).

In the context of financial decisions, Odean (1998), Barber and Odean (1999), and Grinblatt and Keloharju (2009), show that overconfidence leads to "excessive" trading. Statman et al. (2004), show that overconfidence can create a disposition effect, and Ben-David et al. (2008), argue that overconfidence intensifies this effect. Daniel et al. (1998), demonstrate that investor overconfidence (in the "too high precision" sense) may cause excessive volatility. Ben-David et al. (2007), conduct a large survey documenting that managers mis-calibrate their estimates of key financial variables leading them to distorted decisions. Malmendier and Tate (2005a, b), and Malmendier et al. (2007), argue that overconfidence leads managers to prefer using internal financing to external financing, causing the often observed "pecking order of financing." Roll (1986), Hayward and Hambrick (1997), and Malmendier and Tate (2008), show that overconfidence of managers may lead to value-destroying mergers. Landier and Thesmar (2009), show that optimism causes managers to prefer short term to long term financing.

Some researchers also analyze the constructive effects of overconfidence. Cassar and Friedman (2007), show that the initiation of start up activities is positively related to overconfidence. Gervais and Goldstein (2007), argue that overconfidence enhances team productivity. Within a standard principal agent framework, Chen and Chen (2007), show that an agent's effort increases with overconfidence. Oyer and Schaefer (2005), discuss the role of optimism in sorting employees. Bergman and Jenter (2007), analyze the effect of optimism (and other sentiments) on compensation.

This chapter adds to the literature as it analyzes the effect of the overconfidence of effort averse managers on the effort they exert, their compensation and consequently on the welfare of the firm's stakeholders. Our analysis shows that managerial overconfidence causes stockholders to alter the compensation contract they offer to managers and induces managers to exert higher effort levels. It thus helps mitigate the well known conflict of interest between managers and stockholders that induces managers to exert effort levels that are lower than the optimal levels that would be contracted if effort was observable and contractible.⁴ Moreover, the benefits to stockholders from the managers' higher effort levels are larger than the monetary equivalents of the manager's utility losses due to their suboptimal (under the realistic

³For example, when a sample of U.S. students assessed their own driving safety, 82% judged themselves to be in the top 30% of the group (Svenson 1981).

⁴This is consistent with the empirical findings of Puri and Robinson (2007).

distribution) contract and effort level choices. These findings provide an explanation for the persistence of managerial overconfidence.⁵

The paper is constructed as follows. In Sect. 69.2 we present the model. In Sect. 69.3 we describe the analysis method, and in Sect. 69.4 we present the results. In Sect. 69.5 we conclude.

69.2 A Model of Compensation Contracts and the Resulting Managerial Effort

We consider a one period (Holmstrom 1979),-type model where a risk neutral firm employs an overconfident, risk averse and effort averse manager.⁶ We assume that the cash flows, X , of the firm depend on the manager's effort and on exogenous stochastic factors. The manager is assumed to provide some effort which is the minimum necessary to run the firm and hence may be considered observable, but can provide also unobservable extra-effort. The more extra-effort the manager exerts, the higher will be the expected cash flows. Because stockholders cannot observe managers' extra-effort, managerial compensation may depend on the firm's cash flows (which depend on extra-effort), but cannot depend directly on extra-effort.

We assume that the cash flows of the firm, X , are Lognormally distributed with the following distribution function:

$$f(X) = \exp \left\{ -0.5 \left\{ [\log(X) - \mu(Y)] / \sigma \right\}^2 \right\} / \left(X \sigma \sqrt{2\pi} \right) \quad (1)$$

where Y denotes the managerial extra-effort (a managerial choice variable), and $\mu(Y)$ and σ denote, respectively, the mean and the standard deviation of the underlying normal distribution of the natural logarithm of X . We assume that managerial effort increases cash flows, and that overconfident managers overestimate the impact of their effort on cash flows. Formally, we use the following specification for the subjective impact of effort on output:

⁵The persistence of overconfidence has been explained by evolutionary biology. Persistence of overconfidence among managers has been explained by, Gervais and Odean (2001), by the attribution effect. Bernardo and Welch (2001), explain this by an evolutionary process where the overconfident group compensates for its mistakes by its innovative capabilities and hence is able to survive. Goel and Thakor (2008), and KrÅahmer (2003), suggest that overconfidence enhances the chances of managers to succeed in tournaments or contests. Overconfident managers are more willing to take risks and are therefore more likely ending up winning tournament-like contests. Heaton (2002), points out that in the corporate environment, optimistic managers are not likely to be arbitrated away.

⁶In our model we assume symmetry of information between the manager and the firm regarding the distribution of cash flows of the firm except for the different view of the effect of the manager's effort on cash flows.

$$\mu(Y) = \text{Ln}(\mu_0 + 500\lambda Y) - \sigma^2/2 \quad (2)$$

where λ denotes the degree of overconfidence. We assume that stockholders have realistic expectations, which are represented by $\lambda = 1$, and that managers use $\lambda > 1$ to form their expectations. Thus, $f(X)$ can be written as $f(X, \lambda)$, where $f(X, \lambda = 1)$ represents the realistic cash flow distribution while $f(X, \lambda > 1)$ represents the cash flow distribution as viewed by overconfident managers. For notation brevity, we suppress the λ in $f(X, \lambda)$. By the known properties of the Lognormal distribution, the mean and variance of X equal $e^{[\mu(Y)+0.5\sigma^2]}$ and $\left[e^{[2\mu(Y)+\sigma^2]} (e^{\sigma^2} - 1) \right]$, respectively. Thus, it follows from Eq. 2 that a person with a λ overconfidence measure believes that the mean of the cash flows X is: $e^{\mu(Y)+0.5\sigma^2} = \mu_0 + 500\lambda Y$, and that their coefficient of variation is approximately σ .⁷ Since managers and stockholders differ in their perception of the distributions of cash flows, one must be careful in their use. In what follows we refer to the distribution of cash flows as seen by stockholders as the realistic distribution and will make a special note whenever the manager's overconfident beliefs are used.

Except for her overconfidence, the manager is assumed to be rational and to choose her extra-effort so as to maximize the expected value of the following Constant Relative Risk Aversion (CRRA) utility function:

$$U(I, Y) = -I^\alpha(Y_0 - Y)^\beta \quad (3)$$

where I denotes the manager's monetary income, The parameter α is negative, and the utility function represents a constant relative risk aversion of $1 - \alpha$. Y_0 is a parameter that represents the upper bound of the extra-effort level that the manager can invest in the firm. Since β is negative, the higher is Y_0 the lower is the derivative of $U(I, Y)$ with respect to Y . Y_0 can therefore be considered as representing managerial "efficiency" parameter in the sense that the higher it is, the lower is the disutility of extra-effort that results in a given increase in expected cash flows. $Y_0 - Y$ may be thought of as managerial leisure.

Since stockholders cannot observe the manager's extra-efforts they propose compensation schemes that depend on the observed cash flows, but not on Y . Stockholders, which we assume to be risk neutral, strive to make the compensation performance-sensitive in order to better align the manager's incentives with their own. Stockholders offer the manager a compensation package that includes three components: a fixed wage (W) that she will receive regardless of her extra-efforts and of the resulting cash flows, options with a strike price (K) for a fraction (s) of the equity of the firm, and stock grants for another fraction (g) of the firm.

The following timeline of decisions is assumed. At the beginning of the period the firm chooses the parameters of the compensation contract (K , g , W , and s) and offers

⁷More precisely the square of the coefficient of variation is $\left[e^{\sigma^2} - 1 \right]$ which can be approximated by σ^2 since for any small z , $e^z - 1$ is close to z .

this contract to the manager. Observing the contract parameters, and taking into account the effects of her endeavors on firm cash flows and hence on her compensation, the manager determines the extra-effort level Y that maximizes her expected utility. At the end of the period, X is revealed, and the firm distributes the cash flows to the manager and to the stockholders and then dissolves. The priority of payments is as follows. The firm first pays the wages (or only part of them if the cash flows do not suffice). Next, the manager is paid the value of the options, $s\text{Max}[(X - K - W), 0]$. Out of the remaining equity, the manager receives the fraction, g , and the shareholders, the fraction $1-g$.

The manager therefore receives the cash flows $I(X)$ defined by:

$$I(x) = \begin{cases} X \\ (1-g)W + gX \\ (1-g)[W + s\text{Max}(X - W - K, 0)] + gX \end{cases} \text{ when } \begin{cases} X \leq W \\ W \leq X \leq W + K \\ W + K \leq X \end{cases} \quad (4)$$

In the above cash flows formula the first range covers the case where cash flows do not suffice to pay the entire wage. The second range covers the case where the options expire out-of-the-money, and the manager gets the promised wage and a fraction g of the remaining cash flows. The third range represents cash flows large enough so that the options expire in-the-money. In addition to the wage, the manager receives a proportion, s , of the value of the firm above the threshold value of K , and a proportion, g , of the rest. The expected utility of the manager $E\{U[I(X), Y]\}$ which governs her behavior, and her expected compensation $E[I(X)]$, can be obtained by integrating her utility, $U[I(X), Y]$, given in Eq. 3 and her compensation, $I(X)$, given in Eq. 4, respectively. We note that the manager chooses the effort level so as to maximize the expected utility using her perception of the distribution of the firm's final cash flows, while stockholders choose the parameters of the compensation contract using the realistic cash flows distribution to calculate the expected cash flows and managerial compensation.

Shareholders receive all cash flows that are not received by the manager. Since stockholders are risk neutral and rational, stockholders' equity value (SEV) is the expected value of these payments, using the realistic distribution function, and hence⁸:

$$SEV = E(\text{Cashflows}) - E[I(X)] = \int_0^{\infty} Xf(X)dX - E[I(X)] \quad (5)$$

While the derivation of the optimal contract for any set of exogenous parameters is conceptually straightforward, unfortunately, closed form solutions cannot be

⁸Discounting the cash flows by an appropriate risk-adjusted discount rate would yield a linear transformation of equity values. To simplify the presentation, and as is common in the literature, we abstract from that.

obtained in our integrative model. Hence, we resort to simulations to evaluate the optimal contracts and analyze their properties. In addition, the Black-Scholes model cannot be used to evaluate the executive stock options since it takes the values of the underlying asset as given, whereas a crucial aspect of the managerial incentive scheme of our model is that managerial extra-effort and firm value are endogenously determined. We therefore introduce a model that simultaneously simulates the manager's optimal extra-effort level as well as the expected values of the executive stock options, the stock grants the manager receives, and shareholders' equity for each compensation package. We check the robustness of our results by using alternative parameters for the manager's utility function, the impact of managerial extra-effort on firm performance and the distribution functions of the cash flows.

Our model shares many similar features with the Holmstrom (1979), model. The latter however assumes that stockholders have a huge computational capacity and can designate a compensation structure that maximizes the manager's expected utility and their stocks' values, assumption that is quite strong.⁹ To partially circumvent this problem, we employ the following slightly different procedure. We first look at all the possible contracts that may be offered. That is, we consider all possible combinations of K , g , s , and W and then find for each combination the manager's optimal extra-effort, $Y_{K,g,s,W}^*$. We then find for each contract the resulting stockholders equity, the expected managerial compensation and other contract characteristics. We identify the contracts that yield high equity values, and refer to the contracts that are in the top stockholders' equity decile as optimal contracts. Then we consider the impact of managerial overconfidence on stockholders' values that are associated with the optimal and, for robustness, also the other contracts.

69.3 The Simulation Procedures

In calibrating the parameters for the simulations we try to approximately conform to other studies that simulate decisions with effort aversion such as Bitler et al. (2005), and studies that explore the effect of overconfidence on corporate decisions, such as Malmendier and Tate (2005a, b, 2008).¹⁰ Accordingly, we set the parameters in our base case as follows. The coefficient of variation, σ , equals 0.3, and thus the standard deviation is $0.3E(X)$. Since the expected cash flows serve as numeraire, the volatility is determined solely by the coefficient of variation. The expected cash flows as viewed by an overconfident manager with an overconfidence measure of λ are $E(X) = 45,000 + 500\lambda Y$ (i.e., $\mu_0 = 45,000$). The efficiency parameter is $Y_0 = 30$, and the risk aversion and effort aversion parameters are $\alpha = -1$, and $\beta = -0.50$, respectively.

⁹For an extensive criticism of this approach see, e.g., Rubinstein (1998); March (1994); Simon (1957); and Sade et al. (2006).

¹⁰See Appendix A for the explanation for the calibration of our model.

We consider overconfidence levels between $\lambda = 1.5$ and $\lambda = 2.5$. When $\lambda = 1.5$, the manager believes that the contributions of her extra-effort to expected cash flows are 50% higher than the realistic estimates.¹¹ If the manager exerts an effort of $Y = 20$ (this is a round figure close to the average actual effort determined in the base case simulations), she overestimates her contribution to expected cash flows by about 9.1% of total expected cash flows. When $\lambda = 2.5$, the corresponding figure is about 27.3%.

The robustness of the results to deviations from the base case combination of parameters is examined by simulating with several alternative sets of exogenous parameters. We repeat the analysis for many alternative sets of the exogenous parameters: the manager's risk and extra-effort aversion, σ and β , her efficiency, Y_0 , and the volatility measure of cash flows, σ .¹²

We approximate the outcomes of the set of all possible contracts by considering several combinations of several representative values of each of the parameters. We set the upper limits of each of the parameters at such levels that would be high enough so as not to be restrictive. We use integer values between 0% and 5%, in steps of 1%, for the percentage, s , of the firm that is given out as options.¹³ For the percentage of the firm paid out as equity, g , we choose values between 0% and 0.5% in steps of 0.1% and for W we choose values between 2 and 12 in steps of 2. For the representative strike prices we choose values (in 1,000 s) between 15 and 55, in steps of 5. We choose the upper limit of 55 since it is higher than the expected value of the firm without the manager making any extra-effort (45) and close enough to the expected firm value obtainable when the manager exerts the maximal possible effort (60). We do not consider strike prices higher than 55 since out-of-the money options do not turn out optimal in our initial simulations. When the strike price equals zero, an option becomes identical to a stock and hence many of the alternative contract replicate each other (and are less attractive to stockholders when compared with contracts that include options with higher strike prices). In order to make a clear distinction between options and stocks, and because granting options with a very low strike price is very rare, we do not consider contracts with options whose strike prices are close to 0. In selecting the number of alternative values for each variable within the limits we weigh the tradeoff between precision and computational feasibility.

The above choice of representative values for each parameter yields 1944 different contracts for each set of exogenous variables and overconfidence level (6 option percentages $\times$ 6 equity percentages $\times$ 6 wage levels $\times$ 9 strike prices = 1944).

For each contract we then find the manager's optimal extra-effort by solving, for $Y \leq Y_0$, the following problem (using Mathematica):

¹¹ For other overconfidence levels, the belief of the manager is that the contributions of her extra-effort to expected cash flows are $(\lambda - 1)100\%$ higher than the realistic estimates.

¹² Because of scaling there is no need to conduct robustness checks for the expected cash flows.

¹³ Initial simulations indicate that the optimal option awards are likely to be in the range 0–5%, hence we use this range in the present simulations. Similar considerations lead us to the determination of the range for the other variables in our simulations.

$$\text{Max}_Y \int_0^\infty U(I(X), Y) f(X, \lambda) dX \quad (6)$$

where $I(X)$ is given by Eq. 4, $U[I(X), Y]$ by Eq. 3, and $f(X, \lambda)$ by Eq. 1.

Under each overconfidence level, we sort the contracts according to the resulting stockholder equity and identify those that yield the top 10% of stockholder equity (which for brevity we call “optimal”) as well as those in each stockholder equity quintile. We then analyze how overconfidence affects the resulting welfare of the stockholders and the managers assuming that an “optimal” contract is chosen.

69.4 Results and Discussion

In Table 1 we present simulations results that facilitate the examination of the impacts of overconfidence on the effort choice of managers, the expected managerial compensation and the welfare of managers and stockholders. In the five panels of this table we present the resulting managerial effort, the expected value of stockholders equity value (SEV), the expected values of the managerial compensation evaluated using the realistic distribution of cash flows, and the corresponding values when the managers’ optimistic distribution of cash flows is employed. We sort all contracts according to the SEV’s and present averages for six subsets of contracts: the optimal contracts and five contract quintiles by stockholder equity.

In panel A we present the contract outcomes for the optimal contracts and in panels B-E the outcomes for all the contracts. These panels are presented by increasing levels of overconfidence levels from the lowest, $\lambda = 1$, representing a rational manager in panel B, to the most overconfident one with $\lambda = 2.5$ in panel E. The rows in panels B-E correspond to optimal contracts and contract quintiles by decreasing SEV.

We first note that the average SEV’s of the optimal contracts increase as managerial overconfidence increases (see column 2 of Panel A; for example, when $\lambda = 1$ the SEV is 53,347 and when $\lambda = 2.5$ it is 56,589, an increase of about 6%).

Overconfidence increases the willingness of the managers to exert effort because it increases their estimate of the value of the options and stock grants they receive. One observes from column 1 of Table 1 that the average effort levels increase with overconfidence. This is true both for the optimal contracts presented in Panel A and for the other contracts (see column 1 in panels B-E of Table 1). The tendency of the more overconfident managers to work harder for any given set of contract parameters is also gleaned from the following analysis. We examine for all contracts the differences in the effort levels that are chosen by managers with two consecutive (i.e., 1.5 vs. 1, 2 vs. 1.5 and 2.5 vs. 2) levels of overconfidence, holding fixed all the contract parameters. We find that in 98% of the contracts managers with a higher overconfidence level exert more extra-effort.¹⁴

¹⁴Our examination excludes contract pairs where effort levels equal zero for the two consecutive overconfidence levels (about 7.5% of the pairs).

Table 1 Effort levels, Y , stock holders equity values (SEV), and managerial expected compensation according to the objective distribution of cash flows as well as their own subjective distribution. The results for the optimal contracts only are presented in Panel A, and those for all contracts, in panels B–E. Base case: $\alpha = -1$ $\beta = -0.5$, $Y_0 = 30$, $\sigma = 0.3$

	Effort (Y)	Stockholders equity value, SEV	Expected compensation, realistic valuation	Expected compensation, managers subjective valuation
	1	2	3	4
Panel A: Results for the optimal contracts only for all overconfidence levels				
Overconfidence level, λ				
1	18.2	53,347	748.5	748.5
1.5	21.4	55,069	646.7	811.5
2	23.1	55,993	565.9	914.1
2.5	24.2	56,589	494.3	1032.1
Panel B : Results for all contracts, overconfidence level $\lambda = 1$				
Contracts				
Top 10%	18.2	53,347	748.5	748.5
Top quintile	16.1	52,316	757.9	757.9
Second quintile	10.6	49,533	742.8	742.8
Third quintile	5.4	47,122	602.8	602.8
Fourth quintile	0.6	45,037	283.7	283.7
Bottom quintile	0.3	44,735	390.2	390.2
Panel C : Results for all contracts, overconfidence level $\lambda = 1.5$				
Contracts				
Top 10 %	21.4	55,069	646.7	811.5
Top quintile	20	54,318	671.3	832
Second quintile	16	52,296	710.7	840.6
Third quintile	12.8	50,781	638	720.7
Fourth quintile	9	48,673	803.6	863.1
Bottom quintile	1	45,236	259.7	268.1
Panel D : Results for all contracts, overconfidence level $\lambda = 2$				
Contracts				
Top 10 %	23.1	55,993	565.9	914.1
Top quintile	21.9	55,378	580.5	923

(continued)

Table 1 (continued)

	Effort (Y)	Stockholders equity value, SEV	Expected compensation, realistic valuation	Expected compensation, managers subjective valuation
Second quintile	18.7	53,694	679.8	983.9
Third quintile	16	52,324	700.3	918
Fourth quintile	12.4	50,310	900.9	1,072
Bottom quintile	4.4	46,832	371.4	415.3
Panel E : Results for all contracts, overconfidence level $\lambda = 2.5$				
Contracts				
Top 10 %	24.2	56,589	494.3	1032.1
Top quintile	23.1	55,998	540.3	1081.2
Second quintile	20.1	54,389	639.9	1120.6
Third quintile	17.5	53,027	727.2	1086.6
Fourth quintile	14.1	51,075	997	1,314
Bottom quintile	7	48,094	405.4	493.1

The higher penchant of the overconfident manager to exert extra-effort suggests that, compared with employing realistic managers, stockholders can obtain from overconfident managers a given effort level and equity value for a lower-cost (according to stockholders' expectations) compensation contract. Indeed, one observes from column 3 of Panel A in Table 1 that as overconfidence rises, expected managerial compensation declines. The managers' average expected compensation according to the realistic distribution function falls from 748.5, when they are not overconfident to 494.3 when $\lambda = 2.5$, a difference of about 34% (see column 3).¹⁵ However the average expected compensation according to the managers' optimistic distribution of cash flows rises from 748.5 to 1032.1, an increase of about 38% (see column 4). Most of the increase in the SEV accompanying the rise in overconfidence is due to increased effort. For example, out of the increase of 3,242 (= 56,589–53,347) in average SEV that

¹⁵ These expected compensation levels equal, respectively, 1.4% and 0.87% of the corresponding equity values. These ratios are similar to the average of the actual ratios of CEO compensation to company profits (recall that in our one period model stockholders equity corresponds to profit) in a sample of 350 large companies in 2008, based on data from http://www.forbes.com/lists/2009/12/best-boss-09_CEO-Compensation_CompTotDisp.html, and http://money.cnn.com/magazines/fortune/fortune500/2008/full_list.

Table 2 Stakeholders welfare measure SWM (Sum of stockholders equity value, SEV, and managers’ no-effort certainty equivalent measure, NECE) evaluated according to both managers’ and stockholders’ distribution functions for different levels of overconfidence. Base case: $\alpha = -1$ $\beta = -0.5$, $Y_0 = 30$, $\sigma = 0.3$

Overconfidence measure, λ	NECE manager’s evaluation	NECE realistic evaluation	Stockholders equity value, SEV	Stakeholders welfare measure, SWM, managers evaluation	Stakeholders welfare measure, SWM, realistic evaluation
	1	2	3	4	5
1	123.3	123.3	53,347	53,471	53,471
1.5	100.8	64.0	55,069	55,170	55,133
2	94.8	37.4	55,993	56,088	56,030
2.5	94.4	23.9	56,589	56,683	56,613

stockholders enjoy when the overconfidence level grows from 1 to 2.5, only about 8% (254.2/3,242) is due to the savings in expected pay. Most of the rise in the SEV is due to the increase in the effort that the overconfident manager exerts (it rises from $Y = 18.2$, when $\lambda = 1$, to 24.2, when $\lambda = 2.5$, see column 1 of Panel A).

Since it has been shown that as overconfidence rises, so do stockholders’ equity values and the manager’s expected compensation according to her optimistic beliefs, it may seem that overconfidence increases the welfare of the stakeholders of the firm (in this case just stockholders and managers). However, the results show that overconfidence causes the manager to receive lower expected compensation according to the realistic distribution, and it also induces her to work harder than a realistic manager. One may therefore wonder whether overconfidence improves the total welfare of the stakeholders when that welfare is measured according to the realistic distribution function. Stockholders are assumed to be risk neutral, and hence their welfare can be measured by their SEV. Managers on the other hand are risk averse and effort averse, and therefore their expected utility must be converted to some “certainty equivalent” in order to be able to aggregate these measures. Thus, we define a “no-extra-effort and certainty equivalent” (NECE) measure as the amount of certain compensation requiring no extra-effort that the manager would accept in lieu of any given contract with the ascribed effort and compensation. Having converted the welfare of the manager into monetary units, the stakeholders’ welfare is captured by the sum of the NECE and the SEV, which we call the Stakeholders Welfare Measure (SWM).¹⁶

In Table 2 we present the effects of overconfidence on the welfare of the stakeholders: the stockholders and the managers. Since the managers are over-

¹⁶The payments are virtual and outside the model so that incentives and thus efforts would not be affected. This is the widely used Pareto optimality condition for welfare improvement.

Table 3 Sensitivity analysis of effort levels, Y , stock holders equity values (SEV), expected compensation, according to rational stockholders valuation and to managers valuation, and stakeholders' welfare measure, (SWM), to variations in risk aversion α (Panel B), effort aversion β (Panel C), coefficient of variation, σ , (Panel D), and efficiency measure, Y_0 , (Panel E). The base case is presented in Panel A

	Effort (Y)	Stockholders equity value, SEV	Expected compensation, rational stockholders valuation	Expected compensation, managers valuation	Stakeholders welfare measure, rational valuation (SWM)
	1	2	3	4	5
Panel A: Base case $\alpha = -1, \beta = -0.5$					
Overconfidence measure, λ					
1	18.2	53,347	748.5	748.5	53,471
1.5	21.4	55,069	646.7	811.5	55,133
2	23.1	55,993	565.9	914.1	56,030
2.5	24.2	56,589	494.3	1032.1	56,613
Panel B: Higher effort aversion, $\alpha = -1, \beta = -1$					
Overconfidence measure, λ					
1	8.0	48,263	743.0	743.0	48,440
1.5	14.1	51,335	698.3	815.0	51,426
2	17.2	52,945	636.6	913.2	52,998
2.5	19.1	53,973	559.7	1007.5	54,003
Panel C : Lower risk aversion $\alpha = -.5, \beta = -0.5$					
Overconfidence measure, λ					
1	9.4	49,124	572.5	572.5	49,260
1.5	15.1	52,055	509.4	621.9	52,123
2	18.0	53,579	439.1	696.7	53,616
2.5	19.5	54,357	410.4	832.7	54,385
Panel D : Higher standard deviation, $\sigma = 0.4$					
Overconfidence measure, λ					
1	15.9	52,014	927.1	927.1	52,157
1.5	19.7	54,003	845.9	1002.2	54,085
2	21.6	55,052	764.6	1097.9	55,101
2.5	22.8	55,716	681.3	1188.7	55,747
Panel E: Higher efficiency, $Y_0 = 40$					
Overconfidence measure, λ					
1	27.6	57,976	824.3	824.3	58,095
1.5	30.9	59,736	705.3	944.8	59,793
2	32.6	60,685	594.1	1089.0	60,715
2.5	33.4	61,143	558.3	1330.7	61,171

confident we calculate their welfare using both the objective (realistic) distribution function and their subjective distribution. This is done so as to assess their welfare both from their own point of view as well as from society's perspective. Since

stockholders are assumed realistic, their wealth is calculated only using the realistic distribution function.

One observes from columns 1 and 2 of this table that the welfare of the managers decline with overconfidence. This comes as no surprise since it demonstrates that they suffer from their unrealistic view of their talents which leads them to exert too much effort. On the other hand we observe from column 3 that the welfare of the stockholders increases with managerial overconfidence. Moreover, the total welfare measure, SWM, presented in columns 4 and 5 are positively related to overconfidence. This shows that the stockholders' gains from the managers' overexertion of effort and poor choice of contracts exceed the managers' losses. This impact helps explain the persistence of the overconfidence bias as it demonstrates that overconfidence could be welfare enhancing and thus could be a socially desirable character feature.

Table 3 depicts the results of sensitivity analyses that examine the effects of a change in each of the parameters on the behavior of stockholders and managers and on welfare. The analyses provide a robustness check by showing that the results obtained from the base case are qualitatively immune to changes in the parameters. We present the results of only one or two changes in each of the exogenous parameters (α , β , Y_0 , and σ), but we conduct many other simulations, and all provide the same qualitative results.¹⁷

The results in Table 3, demonstrate that SEV, effort, and SWM increase with overconfidence under all the parameter combinations that we use. This shows that our results are robust to a wide range of alternative parameter values.

69.5 Conclusion

Our study addresses an important puzzle in Corporate Finance: Why does the market allow managerial overconfidence to persist in spite of being associated with value-destroying managerial decisions. The current study argues that managerial overconfidence helps mitigate an agency problem whereby realistic managers exert less than the socially optimal level of effort since their effort is not observable and can be compensated only through its impact on a stochastic firm value. Stockholders in this case offer managers a compensation contract that includes bonus stocks and/or options in order to better motivate them. By simulating the compensation contracts that stockholders offer, and the corresponding effort choices of overconfident managers, we demonstrate that overconfident managers exert more effort than realistic ones, thus partially mitigating the above mentioned agency problem. We also construct a measure of the total welfare of the firm's stakeholders (managers and stockholders), taking into account the effort and risk choices of managers. We show that this total welfare measure increases with managerial overconfidence. This indicates that some aspects of managerial overconfidence

¹⁷ These simulations may be obtained from the authors upon request.

may be socially beneficial and may offset their other value destroying effects. These favorable impacts on firm value and total welfare increase the viability of overconfidence in the long run.

Appendix A: Calibration of the Parameters of the Base Case Calculation

We refer to Hek (1999) and Bitler et al. (2005) for the choice of the parameters and the shape of utility functions that depend also on leisure.¹⁸ Hek finds the power utility useful in explaining technological change. Bitler et al. (2005) find that behavior of entrepreneurs can be explained by this type of utility function. We start with our base case $U(I, Y) = -I^\alpha(Y_0 - Y)^\beta$, where $\alpha = -1$ and $\beta = -0.50$. We choose $\alpha = -1$ (a measure of constant relative measure of risk aversion of 2) since this is the parameter that has been widely used in studies of overconfidence: Malmendier and Tate (2005a, b). We examine however also other parameters as robustness checks.¹⁹

We assume that the function $\mu(Y)$ is given by $\mu(Y) = \ln(\mu_0 + Y/2) - \sigma^2/2$. By the known properties of the Lognormal distribution, the mean and variance of this distribution equal $e^{[\mu(Y)+0.5\sigma^2]}$ and $\left[e^{[2\mu(Y)+\sigma^2]}(e^{\sigma^2} - 1)\right]$, respectively. The parameters μ_0 and σ were selected so that, in the absence of managerial extra-effort, the expected cash flows per share are 45,000 and the coefficient of variation (i.e., the ratio of the standard deviation to the expected value) of the cash flow per share is 0.3. Since the expected cash flows serve as numeraire, the ratio of the standard deviation of the cash flows per share to their expected value is a surrogate for the standard deviation of stock returns.

References

- Alpert, M., and Howard Raiffa. 1982. A progress report on the training of probability assessors. In *Judgment under uncertainty: Heuristics and biases*, ed. D. Kahneman, P. Slovic, and A. Tversky. Cambridge: Cambridge University Press.
- Baker, M., R. Ruback, and J. Wurgler. 2006. Behavioral corporate finance: A survey. In *Handbook of corporate finance: Empirical corporate finance*, ed. E. Eckbo, 145–186. Amsterdam: North Holland.
- Barber, Brad M., and Terrance Odean. 1999. The courage of misguided convictions. *Financial Analysts Journal* 6: 41–55.

¹⁸We found additional estimates of effort disutility (leisure utility) in the following papers: Prasch (2001), Kiker and Mendes-de-Oliveira (1990), and Dowell (1985). These estimates varied in the functional form as well as in the level of effort aversion.

¹⁹See, e.g., Bitler et al. (2005), Constantinides et al. (2002), Carpenter (2000), Levy (1994), Epstein and Zin (1991), Mehra and Prescott (1985), and Friend and Blume (1975).

- Barberis, N., and R. Thaler. 2003. A survey of behavioral finance. In *Handbook of the economics of finance: Financial markets and asset pricing*, ed. G. Constantinides, M. Harris, and R. Stulz, 1053–1124. Amsterdam: North Holland.
- Bar-Yosef, Sasson, and Itzhak Venezia. 2008. *An experimental study of overconfidence in accounting numbers predictions*. School of Business, The Hebrew University: Israel.
- Baumann, Andrea O., Raisa B. Deber, and Gail G. Thompson. 1991. Overconfidence among physicians and nurses: The micro-certainty, macro-uncertainty phenomenon. *Social Science & Medicine* 32: 167–174.
- Ben-David, Itzhak, John A. Doukas, and David Hirshleifer. 2008. Is the disposition effect caused by investor loss aversion? Working paper, Ohio State University, USA.
- Ben-David, Itzhak, John R. Graham, and Harvey R. Campbell. 2007. Managerial Overconfidence and Corporate Policies. NBER working paper no. 13711.
- Bergman, Nittai K., and Dirk Jenter. 2007. Employee sentiment and stock option compensation. *Journal of Financial Economics* 84: 667–712.
- Bernardo, Antonio, and Ivo Welch. 2001. On the evolution of overconfidence and entrepreneurs. *Journal of Economics and Management Strategy* 10: 301–330.
- Bitler, M.P., T.J. Moskowitz, and A. Vissing-Jorgensen. 2005. Testing agency theory with entrepreneur extra-effort and wealth. *Journal of Finance* 60 (2): 539–576.
- Camerer, C., and U. Malmendier. 2007. Behavioral economics of organizations. In *Behavioral economics and its applications*, ed. P. Diamond and H. Vartiainen, 235–290. Princeton: Princeton University Press.
- Carpenter, Jennifer N. 2000. Does option compensation increase managerial risk appetite. *Journal of Finance* 55: 2311–2331.
- Cassar, G., and Henry Friedman. 2007. Does overconfidence affect entrepreneurial investment? *Wharton Research Scholars Journal*.
- Chen, Q., and L. Chen. 2007. Study on effect of agent's overconfidence on principal-agent relationship. *Service Systems and Management* 9–11: 1–6.
- Christensen-Szalanski, Jay J., and James B. Bushyhead. 1981. Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance* 7: 928–935.
- Constantinides, G.M., J.M. Donaldson, and R. Mehra. 2002. Junior can't borrow: A new perspective on the equity premium puzzle. *Quarterly Journal of Economics* 117: 269–296.
- Cooper, Arnold C., Carolyn Y. Woo, and William C. Dunkelberg. 1988. Entrepreneurs' perceived chances for success. *Journal of Business Venturing* 3: 97–108.
- Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam. 1998. Investor psychology and security market under- and overreactions. *The Journal of Finance* 53: 1839–1885.
- DeBondt, W.F.M., and R.H. Thaler. 1995. Financial decision-making in markets and firms: A behavioral perspective. In *Handbook in Operations Research and Management Science*, ed. R. Jarrow, V. Maksimovic, and W. Ziemba. Amsterdam: Elsevier/North Holland, 9, 13, 385–410.
- Dowell, Richard. 1985. Risk preference and the work-leisure trade-off. *Economic Inquiry* 23: 691–701.
- Epstein, Larry G., and Stanley E. Zin. 1991. Risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy* 99: 263–286.
- Frank, Jerome D. 1935. Some psychological determinants of the level of aspiration. *American Journal of Psychology* 47: 285–293.
- Friend, I., and M. Blume. 1975. The demand for risky assets. *American Economic Review* 65: 900–927.
- Gervais, Simon, and Itay Goldstein. 2007. The positive effects of biased self-perceptions in firms. *Review of Finance* 11: 453–496.
- Gervais, Simon, and Terrance Odean. 2001. Learning to be overconfident. *The Review of Financial Studies* 14: 1–27.

- Goel, A.M., and A. Thakor. 2008. Overconfidence, CEO selection, and corporate governance. *The Journal of Finance* 63: 2737–2784.
- Grinblatt, Mark, and Matti Keloharju. 2009. Sensation seeking, overconfidence, and trading activity. *The Journal of Finance* 64: 549–578.
- Hayward, M., and D. Hambrick. 1997. Explaining the premiums paid for large acquisitions: evidence of CEO Hubris. *Administrative Science Quarterly* 42: 103–127.
- Heaton, J.B. 2002. Managerial optimism and corporate finance. *Financial Management* 31: 33–45.
- Hek, Paul A. 1999. Endogenous technological change with leisure-dependent utility. *Economic Theory* 14 (3): 669–684.
- Holmstrom, B. 1979. Moral hazard and observability. *Bell Journal of Economics* 10: 74–91.
- Irwin, Francis W. 1953. Stated expectations as functions of probability and desirability of outcomes. *Journal of Personality* 21: 329–335.
- Kidd, John B. 1970. The utilization of subjective probabilities in production planning. *Acta Psychologica* 34: 338–347.
- Kiker, B.F., and M. Mendes-de-Oliveira. 1990. Estimation and valuation of non-leisure time. *Oxford Bulletin of Economics and Statistics* 52: 115–141.
- KrÄahmer, D. 2003. Learning and self-confidence in contests. Discussion paper SPII 2003, 10, Wissenschaftszentrum, Berlin. <http://ssrn.com/abstract=395702>
- Kunda, Ziva. 1987. Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology* 53: 636–647.
- Landier, Augustin, and David Thesmar. 2009. Financial contracting with optimistic entrepreneurs: Theory and evidence. *Review of Financial Studies* 22: 117–150.
- Langer, Ellen J., and Jane Roth. 1975. Heads I Win, Tails it's Chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology* 32: 951–955.
- Levy, Haim. 1994. Absolute and relative risk aversion: An experimental study. *Journal of Risk and Uncertainty* 8: 289–307.
- Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence Phillips. 1982. Calibration of probabilities: The state of the art to 1980. In *Judgment under uncertainty: Heuristics and biases*, ed. Kahneman Daniel, Paul Slovic, and Amos Tversky. Cambridge and New York: Cambridge University Press.
- Malmendier, U., and G. Tate. 2005a. Does overconfidence affect corporate investment? CEO overconfidence measures revisited. *European Financial Management* 11: 649–659.
- . 2005b. CEO overconfidence and corporate investment. *Journal of Finance* 60: 2661–2700.
- Malmendier, U., G. Tate, and J. Yan. 2007. Corporate financial policies with overconfident managers. Working paper, University of California, Berkeley.
- Malmendier, U., and G. Tate. 2008. Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics* 89: 20–43.
- March, James G. 1994. *A primer on decision making: How decisions happen*. New York: The Free Press.
- Marks, Rose. 1951. The effect of probability, desirability, and “Privilege” on the stated expectations of children. *Journal of Personality* 19: 332–351.
- Mehra, R., and E.C. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15: 145–161.
- Neale, Margaret A., and Max H. Bazerman. 1990. *Cognition and rationality in negotiation*. New York: The Free Press.
- Odean, Terrance. 1998. Volume, volatility, price and profit when all trades are above average. *Journal of Finance* 53: 1887–1934.
- Oyer, Paul, and Scott Schaefer. 2005. Why do some firms give stock options to all employees? An empirical examination of alternative theories. *Journal of Financial Economics* 76: 99–133.
- Prasch, Robert E. 2001. Work, leisure, and the labor supply curve: A reply to David Spencer. *Journal of Economic Issues* 35: 1001–1007.

- Puri, Manju, and David T. Robinson. 2007. Optimism and economic choice. *Journal of Financial Economics* 86: 71–99.
- Roll, R. 1986. The Hubris hypothesis of corporate takeovers. *Journal of Business* 59: 197–216.
- Rubinstein, Ariel. 1998. *1998. Modeling bounded rationality*: MIT Press.
- Russo, J. Edward, and Paul J.H. Schoemaker. 1992. Managing overconfidence. *Sloan Management Review* 33: 7–17.
- Sade, Orly, Charles Schnitzlein, and Jaime F. Zender. 2006. Competition and cooperation in divisible good auctions: An experimental examination. *Review of Financial Studies* 19: 195–235.
- Shefrin, H. 2000. *Beyond greed and fear: Understanding behavioral finance and the psychology of investing*. Cambridge, MA: Harvard Business School Press.
- Simon, Herbert. 1957. A behavioral model of rational choice. In *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting*. New York: Wiley.
- Staël von Holstein, and S. Carl-Axel. 1972. Probabilistic forecasting: An experiment related to the stock market. *Organizational Behavior and Human Performance* 8: 139–158.
- Statman, Meir, Steven Thorley, and Keith Vorkink. 2004. Investor overconfidence and trading volume. Working paper, University of Santa Clara.
- Svenson, Ola. 1981. Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica* 47: 143–148.
- Taylor, Shelley, and Johathon D. Brown. 1988. Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin* 103: 193–210.
- Wagenaar, Willem, and Gideon B. Keren. 1986. Does the expert know? The reliability of predictions and confidence ratings of experts. In *Intelligent decision support in process environments*, ed. Hollnagel Erik, Giuseppe Mancini, and David D. Woods. Berlin: Springer.
- Weinstein, N. 1980. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology* 39: 806–820.
- Weinstein, N., and W. Klein. 2002. Resistance of personal risk perceptions to debiasing interventions. In *Heuristics and biases: The psychology of intuitive judgment*, ed. T. Gilovich, D. Griffin, and D. Kahneman. Cambridge, UK: Cambridge University Press.

Current Versus Permanent Earnings for Estimating Alternative Dividend Payment Behavioral Model: Theory, Methods and Applications

70

Cheng-Few Lee, Hong-Yi Chen, Alice C. Lee, and Yuhsin Tai

Abstract

The main purposes of this chapter are to: (1) theoretically explain why firms generally allocate permanent earnings and transitory earnings between dividends payments and retained earnings; (2) develop alternative methods for decomposing current earnings into permanent and transitory components; (3) empirically estimate alternative dividend payment behavior models by using two alternative permanent EPS estimates for both individual firms and pooled data; and (4) test Lambrecht and Myer's (2012) theoretically results related to alternative dividend payment behavior models. We find that the average long-term payout ratio is downward biased and the average estimated intercept is generally upward biased when current instead of permanent EPS are used. We also find that the combined model performs well to deal with both measurement errors and specification errors in describing the dividend payment behavior model.

C.-F. Lee (✉)

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: cflee@business.rutgers.edu

H.-Y. Chen

Department of Finance, National Chengchi University, Taipei, Taiwan

e-mail: fnhchen@nccu.edu.tw

A. C. Lee

Center for PBBEF Research, Phoenix, AZ, USA

e-mail: alice.finance@gmail.com

Y. Tai

Department of Finance and Economics, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

e-mail: tzutai30@gmail.com

Keywords

Current earnings · Current EPS · Permanent earnings · Permanent EPS · Dividend behavior models · Specification analysis · Partial adjustment coefficient · Long-term payout ratio

70.1 Introduction

Earnings of a firm are allocated to retained earnings or dividend payments by a financial decision. Retained earnings are internal sources of funds that provide additional financial capital for either the expansion of the firm or a financial reserve against future contingencies. Dividends are generally distributed to stockholders to satisfy their need for liquidity or other uses according to their preference functions. It is well known that earnings of a firm can be classified into either permanent or transitory components. Permanent earning power creates the permanent component, and the transitory component is composed of income of a temporary nature. Modigliani and Miller (1958, 1961, 1963, 1966) have argued that a firm's market value is determined by its permanent (expected) earnings, not transitory components of income.

The transitory component of a firm's earnings originates from a temporary change in market conditions, a temporary change in accounting method, or any other nonpermanent change that would cause earnings to fluctuate over time. Latane and Jones (1979) discuss the importance of unexpected earnings of firms as signaling information in financial management and investment analysis. However, to the best of our knowledge, no acceptable method for decomposing current earnings into permanent (expected) and transitory (unexpected) earnings has been previously developed.

In addition, forecasts of dividends are important to both security analysts and financial managers, and either conditional or unconditional methods are generally used to forecast dividend payments. The most popular conditional dividend-forecasting models are the partial adjustment model developed by Lintner (1956) and the information content model discussed by Ang (1975); several others are also available.

Lintner (1956) uses survey data to develop the dividend payment behavior model describing how managers determine their dividend payment. Lee et al. (1987) use a partial adjustment and adaptive expectation model to generalize Lintner's dividend behavior model. Since then, Lintner's model has been widely used in finance research, such as Marsh and Merton (1987), Lee and Primeaux (1991), Garrett and Priestley (2000), and Lambrecht and Myers (2012). Miller and Modigliani (1966) show that current earnings used to estimate cost of equity capital is subject to measurement error problem. Therefore, using current EPS to estimate Lintner's dividend behavior model might be also subject to measurement

error problem. Kumar and Lee (2001) proposed a coarse signaling model to find the determinants of dividend smoothing. The empirical studies illustrated that dividend smoothing is associated with not only permanent earnings but also other risk factors.

Marsh and Merton (1987) have theoretically developed an aggregate dividend behavior model and empirically used S&P 500 index as proxy to measure aggregate permanent earnings. However, they did not explicitly develop a method estimate permanent earnings. Garrett and Priestley (2000) have generalized the Marsh and Merton model by including both the S&P 500 index and permanent earnings in their dividend payment behavior model. In addition, they proposed a common Kalman filter approach to estimate aggregate permanent earnings. Pan (2001) investigated the aggregate dividend behavior of US companies during 1871–1993. The empirical results indicated that the change of dividend is related to the change of permanent earnings and has the revision feature to a long-term dividend payout target. Ham, Kaplan, and Leary (2020) showed that dividend changes convey the information of the future earnings that persists for at least three years. The evidence supported that the dividend change is positively proportional to the level of permanent earnings.

To the best of our knowledge, Lee and Primeaux (1991) is the first paper that empirically shows how current EPS can be decomposed into permanent and transitory EPS. In addition, they used permanent instead of current EPS to estimate Lintner's dividend payment behavior model for individual companies. Most recently, Lambrecht and Myers (2012) have theoretically shown that permanent instead of current EPS should be used to estimate the dividend payment behavior model for individual companies. They also provide specification analysis to show how the dividend payment behavior model can be misspecified if current EPS is applied.

The main purposes of this chapter are to: (1) theoretically explain why firms generally allocate permanent earnings and transitory earnings between dividends payments and retained earnings; (2) develop alternative methods for decomposing current earnings and dividends into permanent and transitory components; (3) empirically estimate alternative dividend payment behavior models by using two alternative permanent EPS estimates for both individual firms and pooled data; and (4) test Lambrecht and Myer's (2012) theoretically results related to alternative dividend payment behavior models.

Section 70.1 is the introduction. Section 70.2 discusses theoretical determination of firm's permanent and transitory earnings and dividends. The relationship between accounting earnings and economic earnings is also discussed. Section 70.3 discusses alternative models for decomposing current earnings and dividends into permanent and transitory components, according to methods proposed by Darby (1972, 1974), Lee and Primeaux (1991), and Garrett and Priestley (2000). In Sect. 70.4, empirical results of testing model discussed in Sect. 70.3 are revealed in terms of individual firms and pooled data. We also perform the

empirical tests of Lambrecht and Myer's (2012) theoretical results of permanent EPS and their specification analysis of dividend payment behavior model in terms of current EPS. Sect. 70.5 provides a summary and some concluding remarks.

70.2 Theoretical Determination of Firm's Permanent and Transitory Earnings and Dividends

In the evolution of the consumption function, which is one of the key concepts in Keynesian economics, several important theories were developed to explain how consumers adjust consumption expenditures to accommodate changes in their levels of income. One of these theories is the permanent income hypothesis developed by Friedman (1957).¹

The permanent income hypothesis shows that consumption is not a function of current income but a function of permanent income. Total income is composed of two components permanent income and transitory income. Transitory income is not fully anticipated and it may be positive or negative. That is, a prize would constitute a positive transitory income component while a loss of income from temporary illness or layoff would constitute a negative component of transitory income. Friedman explains that these transitory elements would not affect consumption expenditures.

The permanent income hypothesis is applied to the finance theory, and a new theory of dividend payments by business can be developed. The income of interest here is the income of the business firm and dividends are analogous to consumer consumption expenditures.

The level of permanent income earned by a firm determines the permanent dividends it can pay out to stockholders. Permanent income is essentially an average of current and past earnings of the firm. Current income, therefore, can be divided into two components:

$$E = E^P + E^T \quad (1)$$

where E is the current income per share of the firm, E^P is the permanent income per share of the firm, and E^T is the transitory income per share of the firm.

Transitory income may be positive or negative, and current income will differ from permanent income by the amount of transitory income. A business earns transitory income, which is really unanticipated earnings, from windfall profits from any sources. For example, oil companies earn transitory income from the increased prices they received from selling products made from crude oil produced domestically. Firms incur negative transitory income if they experience an uninsured

¹When Friedman received the Nobel prize in economics, this work was cited as one of his major contributions.

catastrophic event such as the destruction of a plant by a disaster of any kind or an unexpected strike by employees. The transitory components of income, positive and negative, should cancel out over the permanent income time horizon. Transitory components, however, are always present during shorter time periods.

Eisner (1967, 1978) developed a permanent income theory for investment decisions. If firm investment essentially depends upon internal sources of funds, the nature of retained earnings is an important factor affecting the decision to undertake long-term or short-term investment.

Retained earnings can conceptually be decomposed into two components, permanent and transitory. Dividends can also be divided into two similar components:

$$D = D^P + D^T \quad (2)$$

where D is the current dividends per share paid by the firm, D^P is the permanent dividends per share paid by the firm, and D^T is the transitory dividends per share paid by the firm.

Permanent dividends are only one component of dividends, and total dividends may be larger than permanent dividends, depending upon the level of transitory dividends. Permanent dividends are dividends that the business firm systematically pays based on its permanent earnings, dividends paid out of transitory earnings would constitute extra dividends.

Weston et al. (2004) and others generally explain that a firm may have one of three dividend policies: (1) stable dollar amount per share, (2) constant payout ratio, or (3) a compromise – lower regular dividend, plus extras. No matter what policy is used, all income is either paid out in dividends or retained by the business in the form of retained earnings:

$$E - (D^P + D^T) - R = 0 \quad (3)$$

where R is the retained earnings per share of the firm.

Transitory dividends are paid from transitory income and are short run in nature. They are part of the short-run measure of dividend yield. In contrast, permanent dividends are paid from permanent earnings, are long run in nature, and constitute all of the long-run measure of dividend yield. Miller and Scholes (1982) have demonstrated that short-run and long-run dividend yield each have different implications in testing the effectiveness of alternative dividend policies on security rate of return determination. Our theoretical framework, decomposing income and dividend payout into permanent and transitory components, elaborates upon their theoretical justification of short- and long-run dividend yield measurements. Generally, transitory earnings are not used for payment of permanent dividends. However, transitory dividends can come from either transitory or permanent earnings.

Different sources of dividend payment (i.e., permanent or current income) may have different implications in determining a firm's dividend payment behavior. This condition provides the motivation for examining both permanent and current earnings per share for describing a firm's dividend payment behavior in the empirical

section of this work. In the next section, we will discuss alternative methods for decomposing current EPS into permanent and transitory EPS components.

70.3 Alternative Methods for Decomposing Current EPS into Permanent and Transitory EPS Components

In this section, we will discuss four alternative methods to decompose current EPS into permanent EPS and transitory EPS components. These four methods are: (1) Darby's (1974) method, (2) Lee and Primeaux's (1991) method, (3) Garrett and Priestley's (2000) Kalman filter method, and (4) Lambrecht and Myer's (2012) method.

70.3.1 Darby's (1974) Method

We follow Darby's (1974) method to decompose current EPS into permanent and transitory EPS components. Theoretically, the relationship between current dividend and permanent earning can be defined as

$$D_{i,t} = \alpha + \beta E_{i,t}^P + \varepsilon_{i,t} \quad (4)$$

where $D_{i,t}$ and $E_{i,t}^P$ are current dividends and permanent earnings per share for i^{th} firm in period t , respectively. In addition, $\varepsilon_{i,t}$ is a random variable with mean zero and variance σ_ε^2 . Since $E_{i,t}^P$ is not directly observable, we assume that current expectations are derived by modifying permanent expectations in light of current experience. That is,

$$E_{i,t}^P = (1 - \lambda_i)E_{i,t} + \lambda_i(1 + C)E_{i,t-1}^P, \quad 0 \leq \lambda_i < 1 \quad (5)$$

where λ_i represents the weight used to calculate the permanent EPS and C represents the trend rate of EPS growth.

According to Darby (1974), the initial value of permanent EPS $E_{i,0}^P$ and trend rate C can be derived from estimating the EPS trend regression

$$\ln E_{i,t} = a_1 + a_2 t + u_t \quad (6)$$

where u_t is the error term.

After a_1 and a_2 are estimated, $E_{i,0}^P$ and C can be defined as

$$E_{i,0}^P = e^{\hat{a}_1}, \quad \log(1 + C) = \hat{a}_2 \quad (7)$$

To estimate the optimal weights λ_i , we first substitute estimated $E_{i,0}^P$ and $C = e^{\hat{a}_2} - 1$ into Eq. (5) to compute alternative $E_{i,t}^P$ series for $\lambda_i = 0, x, 2x, 3x, \dots$,

1 where x is the interval of estimate for λ_i that either minimize sum of squared residuals or maximize adjusted R -squared of Eq. (4).

70.3.2 Lee and Primeaux's (1991) Method

Fama and Babiak (1968), Kmenta (1986), and Lee et al. (1987) propose the adaptive expectation model to determine the permanent EPS, $E^P_{i,t}$ as

$$E^P_{i,t} - E^P_{i,t-1} = (1 - \lambda_i)(E_{i,t-1} - E^P_{i,t-1}) \quad (8)$$

Equation (8) can be rewritten as:

$$E^P_{i,t} = (1 - \lambda_i)E_{i,t} + \lambda_i E^P_{i,t-1}, \quad 0 \leq \lambda_i < 1 \quad (9)$$

By Koyck transformation, Kmenta (1986) shows that Eqs. (4) and (9) can derive:

$$D_{i,t} = \alpha_0 + \beta_0 E_{i,t} + \gamma D_{i,t-1} \quad (10)$$

where $\alpha_0 = \alpha(1 - \lambda_i)$, $\beta_0 = \beta(1 - \lambda_i)$, $\gamma = \lambda_i$. If λ_i approaches zero, then $E^P_{i,t} = E_{i,t}$. This implies that the permanent EPS is equivalent to the current EPS.

By comparing Eq. (10) to Eq. (5), it is obvious that Eq. (5) is a reduced form of Eq. (10) if C is equal to zero.

To empirically estimate the permanent EPS defined in Eq. (9), we can run the regression and obtain the estimated λ_i which is equal to estimated γ . Using estimated λ_i , current EPS, and initial permanent EPS described in Darby's method in Sect. 70.3.1, we can estimate permanent EPS in period t .

70.3.3 Garrett and Priestley's (2000) Kalman Filter Method

Following Garrett and Priestley's (2000) method, we define the relationship among current EPS, $E_{i,t}$, permanent EPS, $E^P_{i,t}$, and transitory EPS, $E^T_{i,t}$, as follows:

$$E_{i,t} = E^P_{i,t} + E^T_{i,t} \quad (11)$$

To complete the model, we need to specify equation that governs the evolution of the unobservable permanent EPS:

$$E^P_{i,t} = E^P_{i,t-1} + \beta_{t-1} + v_t \quad (12)$$

$$\beta_t = \beta_{t-1} + \eta_t \quad (13)$$

where the permanent EPS, $E^P_{i,t}$, evolves as a random walk with a changing trend, β_t . To extract a measure of permanent EPS, we treat measurement Eq. (11) and

transition Eqs. (12) and (13) as defining an unobserved components model and estimate it via the Kalman filter.

70.3.4 Lambrecht and Myer's (2012) Method

Using the joint determination of manager's rent and cash dividend payment to equity holders, Lambrecht and Myers (2012) derive a Lintner dividend payment behavior in terms of permanent income as:

$$d_t = a_0 + a_1 d_{t-1} + Y_t + e_t, \quad (14)$$

where d_t and d_{t-1} are total dividend payout at time t and $t-1$, respectively; Y_t is the firm's permanent income at time t . Lambrecht and Myers (2012) argue that permanent income Y_t is not observable but theoretically could be estimated from current operating profit and the market's expectation of future profits.

They define permanent income Y_t as the rate of return on the sum of current income and the present value of all future income, net of debt service, but before rents. It is an annuity payment that, given expectations at time t , could be sustained forever: $Y_t = \frac{\rho}{1+\rho-\mu} (K^0 \pi_t - (1+\rho-\mu)D_{t-1}) = \frac{\rho}{1+\rho-\mu} (K^0 \pi_t - (1+\rho-\mu)D_{t-1})$. If the profit margin π_t follows the autoregressive process $\pi_t = \mu \pi_{t-1} + \eta_t$, then permanent income Y_t can be simplified as²:

$$Y_{i,t} = \frac{\rho_i}{1+\rho_i-\mu_i} (K_i^\phi \pi_{i,t} - (1+\rho_i-\mu_i)TD_{i,t-1}), \quad (15)$$

where $K_i^\phi \pi_{i,t}$ is total operating income without corporate tax for i^{th} firm in period t ; $TD_{i,t-1}$ is the total debt for i^{th} firm in period $t-1$; ρ_i is interest rate; and μ_i is the autoregression coefficient for operating income of the firm i . In the limiting case where π_t follows a random walk ($\mu = 1$), permanent income approaches $K^0 \pi_t - \rho TD_{t-1}$, that is, current net income measured before rents but after interest.

Lambrecht and Myers (2012) have briefly discussed how corporate tax can affect permanent income defined in Eq. (15); however, they did not develop a closed-form solution for permanent income with corporate tax. Therefore, their permanent income defined in Eq. (15) does not exactly follow the concept of either economics or accounting.

Lambrecht and Myers (2012) claim that the Lintner model as traditionally estimated can be defined as

$$\Delta d_t = b_0 + b_1 TE_t + b_2 d_{t-1} + u_t, \quad (16)$$

where $\Delta d_t = d_t - d_{t-1}$; the current reported earnings is $TE_t \equiv p_t + \tau_t - \rho TD_{t-1}$; p_t and τ_t are permanent and transitory components, respectively. ρTD_{t-1} is the

²See Appendix A for the detailed definition of permanent income and its related implications.

component neither permanent nor transitory component of earnings. The coefficient b_2 on lagged payouts is interpreted as (the negative of) the speed of adjustment (SOA) and the coefficient b_1 on earnings as the product of the long-term payout ratio and the SOA.

Under their definition of TE_t , ρTD_{t-1} is the most important term in obtaining the true model.

According to Lambrecht and Myers (2012), the true model is:

$$\Delta d_t = \kappa + \frac{\rho\beta\alpha SOA}{1 - \beta\mu} TE_t - SOA d_{t-1} - \frac{\rho\beta^2\mu\alpha SOA}{1 - \beta\mu} \tau_t - \frac{\rho\beta(1 - \mu)\alpha SOA}{1 - \beta\mu} TD_{t-1} + e_t \quad (17)$$

where κ is the constant term of dividend behavior model, it is generally used to measure the degree of reluctance to cut dividend, α is defined as percentage of earnings paid as cash dividend, $\beta = 1/(1 + \rho)$.

Therefore, the estimates for the coefficients from Eq. (16) will be biased and inconsistent unless the omitted variables TD_{t-1} and τ_t are orthogonal to the included variables (Greene (1993), p. 246). The omitted variables are likely to be correlated with the included variables, given the definition of the earnings variable TE_t and because d_{t-1} is linked with TD_{t-1} through the budget constraint. The variance of the estimates and of the error terms are also biased. Thus, the usual confidence interval and hypothesis testing procedures can give misleading conclusions about statistical significance.

In practice, however, the misspecification of the traditional Lintner model in Eq. (16) may not be all that severe. First, corporate earnings or cash flows are highly persistent for mature, stable companies with low earnings volatility (see Dichev and Tang [2009] and Frankel and Litov [2009]). As $\mu \rightarrow 1$ the term in TD_{t-1} in Eq. (18) vanishes and the omitted variable problem with respect to TD_{t-1} disappears. Second, the transitory income component τ_t may account for only a small part of the total earnings TE_t of a mature company. Thus, the correlation between τ_t and TE_t may be small too. In other words, current earnings TE_t may be highly correlated with permanent income when transitory income is small. Of course, TE_t becomes a noisy measure of permanent income when transitory income is volatile and important. The traditional Lintner regression Eq. (16) may therefore give quite different results from the model specified in Eq. (17).

If μ approaches to 1, then the problem associated with τ_t can be resolved by using Darby's approach to calculate permanent earnings. If μ does not approach to 1, then Eq. (17) can be modified as

$$\Delta d_t = b_0 + b_1 TE_t^P + b_2 d_{t-1} + b_3 \rho TD_{t-1} + u_t, \quad (18)$$

where TE_t^P is the estimated permanent earnings in terms of Eq. (5).

Equation (18) is obtained by combining Lambrecht and Myer's (2012) theory and Darby's method of estimating permanent earnings. This specification solves both specification errors and the transitory components of earnings. This new specification is the most important contribution of this research.

Darby's method is relied upon optimal R-square searching for optimal λ_i , while Lee and Primeaux's method relies only regression coefficient estimates. Therefore, Lee and Primeaux's method is empirically easier to estimate permanent EPS. We will use both methods to estimate permanent EPS in the next section. Garrett and Priestley's (2000) Kalman filter method is relatively restrictive in estimating permanent EPS. Therefore, we will use only Darby's method, Lee and Primeaux's method, Lambrecht and Myer's method, and the new method by combining Darby's method and Lambrecht and Myer's method which is defined in Eq. (18) for empirical investigation in next section.

70.4 Empirical Results in Estimating Two Alternative Dividend Behavior Models

In this section, we use EPS and DPS data of 608 firms from Compustat, which has at least 30 years consecutive data by 2011, to perform these empirical studies. The empirical studies include: (1) Darby's method and Lee and Primeaux's method, (2) Lambrecht and Myer's method, and (3) combined model as defined in Eq. (18). EPS, DPS, and payout ratio information for 608 firms are presented in Appendix C following the descending order of payout ratios.

In Appendix C, there are 605 firms with positive payout ratios which are smaller than one. The payout ratios of Weyerhaeuser Co and Rexam Plc are 1.0493 and 1.0205, respectively. The payout ratio for Weyerhaeuser Co is larger than one because of paying special dividend \$405 million in 2010. The earnings per shares for Rexam Plc are -0.83 , -2.37 , -0.85 , and -0.29 in 1996, 2002, 2003, and 2009 respectively. However, this company paid dividends per shares 0.2799, 1.3482, 3.8125, and 3.3558 for these four years. This is the main reason that this company obtained an average payout ratio (1.0205) above one. The payout ratio of Dart Group Corp, which is listed in the last firm in Appendix C, is -6.5141 . The earnings per shares of Dart Group Corp are -10.96 , -4.1 , -39.57 , -7.88 , -8.73 , -19.81 in 1987, 1993, 1994, 1995, 1996, and 1997, respectively. However, this company uses a constant dividend payout (0.1332) during 1972–1997. Therefore, the average EPS and DPS are -0.02 and 0.1303, respectively, and average payout ratio for this company is -6.5141 . It is worthwhile to know that this company bankrupts in 1998. The Appendix C shows that average EPS, DPS, and payout ratio are 2.4290, 0.9159, and 0.3636, respectively. The standard deviation for EPS, DPS, and payout ratio are 1.8977, 0.3768, and 0.0985, respectively. The skewness for EPS, DPS, and payout ratio are 3.3799, 1.7729, and -17.2978 , respectively. In addition, the kurtosis for EPS, DPS, and payout ratio are 20.6265, 4.7306, and 380.6515, respectively. From these statistics of EPS, DPS, and payout ratio, we conclude that the statistical distributions of these three variables are not normally distributed. Therefore, using the pooled data to perform regressions might result in problems with testing the significant estimated coefficients of regression. Hence, we believe that using individual firms' data to estimate dividend behavior model can give more information than pooled EPS and DPS data. Therefore, in this section, we use both individual firms' data and pooled data to perform empirical studies.

70.4.1 Darby's Method and Lee and Primeaux's Method

70.4.1.1 Results from 608 Individual Regressions

In this section, we will use current and permanent EPS measures to estimate following two alternative dividend payment behavior models as:

$$D_{i,t} = c_0 + c_1 E_{i,t} + c_2 D_{i,t-1} + u_{i,t} \quad (19a)$$

$$D_{i,t} = c_0 + c_1 E_{i,t}^P + c_2 D_{i,t-1} + u_{i,t} \quad (19b)$$

$$D_{i,t} = c'_0 + c'_1 E_{i,t} + c'_2 D_{i,t-1} + c'_3 D_{i,t-2} + u'_{i,t} \quad (20a)$$

$$D_{i,t} = c'_0 + c'_1 E_{i,t}^P + c'_2 D_{i,t-1} + c'_3 D_{i,t-2} + u'_{i,t} \quad (20b)$$

Following Eq. (11), the current EPS, $E_{i,t}$, can be decomposed into permanent EPS, $E_{i,t}^P$, and transitory EPS, $E_{i,t}^T$. If we use Eq. (19a) instead of Eq. (19b) to estimate c_1 , the estimated c_1 will be subject to errors-in-variable problem and the estimated c_1 will be downward biased. Following Lee et al. (2015, 2019), we have analyzed the impact of this kind of errors-in-variable problem in Appendix B in details. We now analyze the biased associated with estimated c_1 and c_2 as follows:

Case 1: Under the assumption that $COV(E_{i,t}^P, D_{i,t-1}) = 0$, we can follow Eq. (B10) in Appendix B to obtain the biased associated with estimated c_1 and c_2 as follows:

$$\begin{aligned} \text{plim } \hat{c}_1 - c_1 &= \frac{-c_1 \sigma_1^2}{(\sigma_{E_{i,t}^P}^2 + \sigma_1^2)} \text{ and } \text{plim } \hat{c}_2 - c_2 = \frac{\sigma_1^2 (\sigma_{D_{i,t-1}}^2 - c_2 \sigma_{D_{i,t-1}}^2)}{\sigma_{D_{i,t-1}}^2 (\sigma_{E_{i,t}^P}^2 + \sigma_1^2)} \\ &= 0 \end{aligned} \quad (21a)$$

where σ_1^2 is the variance of $E_{i,t}^T$.

Case 2: Under the assumption that $COV(E_{i,t}^P, D_{i,t-1}) \neq 0$, we can follow Eq. (B13) to obtain the biased associated with estimated c_1 and c_2 as follows:

$$\begin{aligned} \text{plim } \hat{c}_1 - c_1 &= \frac{-c_1 \sigma_1^2}{\sigma_{E_{i,t}^P}^2 - b_{D_{i,t-1}E_{i,t}^P} + \sigma_1^2} \text{ and } \text{plim } \hat{c}_2 - c_2 \\ &= c_1 b_{D_{i,t-1}E_{i,t}^P} \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_{E_{i,t}^P}^2 (1 - R_{E_{i,t}^P D_{i,t-1}}^2)} \right) \end{aligned} \quad (21b)$$

where $b_{D_{i,t-1}E_{i,t}^P}$ is the auxiliary regression coefficient of a regressing $D_{i,t-1}$ on $E_{i,t}^P$, and $R_{E_{i,t}^P D_{i,t-1}}^2$ is the correlation coefficient between $E_{i,t}^P$ and $D_{i,t-1}$.

Equations (21a) and (21b) imply that the estimated c_1 are downward biased. Therefore, the estimated intercept $\hat{c}_0$ as defined in Eq. (21c) is upward biased.

$$\hat{c}_0 = \bar{D}_{i,t} - \hat{c}_1 \bar{E}_{i,t} - \hat{c}_2 \bar{D}_{i,t-1} \quad (21c)$$

Therefore, we need to deal with this kind of errors-in-variable problem.

First, we will use Darby's method to estimate permanent EPS as defined in Eqs. (4), (5), (6), and (7), and use Lee and Primeaux's method to estimate permanent EPS as defined in Eq. (10). We then use DPS and both current EPS and permanent EPS to estimate Eqs. (19a), (19b), (20a), and (20b). From the optimal search of λ_i by Darby's method, we estimate λ_i for 608 firms and found that there are 153 estimated λ_i equal to one and 45 estimated λ_i equal to zero. The estimated λ_i for other 410 firms are between 0 and 1. By using Lee and Primeaux's method, we find that there are 580 estimates of λ_i either larger than zero or less than one. The other 28 estimates of λ_i are equal to zero.

From the regression results of Eqs. (19a) and (19b), we calculated the averages of the estimated intercept, the estimated C_1 and the estimated C_2 , and their results are presented in columns 1, 2, and 3 of Table 1A. Similarly, the averages of the estimated intercept, the estimated C_1 , the estimated C_2 , and the estimated C_3 in Eqs. (20a) and (20b), can be found in columns 4, 5, and 6 of Table 1A. By comparing the average estimated C_1 of Eqs. (19a) and (19b) presented in Table 1A, we found that the average estimates of C_1 associated with permanent EPS calculated by both Darby's method and Lee and Primeaux's method are significantly higher than those estimates associated with current EPS. Miller and Modigliani (1966) have shown that there exists errors-in-variable problem if the current earnings instead of permanent earnings are used to estimate regression coefficient. Therefore, the regression coefficients associated with current EPS instead of permanent EPS are subject to errors-in-variable problem as presented in Eqs. (21a) and (21b). In addition, Almeida et al. (2010) have used investment equations to show how measurement error can affect the estimated regression coefficients for investment equations. Following the explanation in Eq. (21c), we found that the average intercept from Eq. (19a) is significantly larger than that of Eq. (19b) by using Darby's method.

From columns 4, 5, and 6 of Table 1A, we found that there are 9.7%, 10.86%, and 10.36% of estimated C_3 significantly different from zero at 5% significant level. This implies that there exists specification error in original Lintner model for some companies.

Table 1B presents the distribution information of partial adjustment coefficients and long-term payout ratio for 608 firms. We found that these two parameters have a skewed distributed with six outliers. To deal with this problem, we calculate median and trimmed average for both average partial adjustment coefficients and long-term payout ratio.

The outliers of long-term payout ratios in Eq. (19b) are -463.0488 , -101.9549 , -70.1538 , 43.4411 , 48.6478 , and 50.2931 for companies G & K Services Inc., Automatic Data Processing Inc., Stepan Co., Echlin Inc., Goodrich Corp., and Marathon Oil Corp., respectively.

Table 1A Individual regression results for Eqs. (19a), (19b), (20a), and (20b)

Dependent variable	Eq. (19a)	Eq. (19b)	Eq. (19b)*	Eq. (20a)	Eq. (20b)	Eq. (20b)*
	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$
Intercept	0.1350 (0.3711)	-0.0045 (2.9668)	0.1354 (0.5354)	0.1614 (0.4362)	-0.0329 (3.1519)	0.1158 (0.7489)
Median	0.0794 24.84%	0.0503 36.35%	0.1049 34.05%	0.0975 23.68%	0.0505 31.09%	0.0970 28.62%
$E_{i,t}$	0.0977 (0.1365)			0.0764 (0.1108)		
Median	0.0719 61.02%			0.0482 52.96%		
$E^P_{i,t}$		0.2568 (2.2324)	0.1215 (0.2947)		0.2833 (2.4786)	0.1286 (0.3860)
Median		0.1607 66.78%	0.0829 48.85%		0.1642 62.01%	0.0847 47.53%
$D_{i,t-1}$	0.5764 (0.2566)	0.4634 (0.2607)	0.5420 (0.2638)	0.6238 (0.3094)	0.5364 (0.3090)	0.5861 (0.3128)
Median	0.6253 87.99%	0.5057 75.33%	0.5797 82.07%	0.6590 84.05%	0.5659 73.85%	0.6282 76.48%
$D_{i,t-2}$				-0.0027 (0.3154)	-0.0879 (0.2198)	-0.0377 (0.3199)
Median				-0.0071 9.70%	-0.0875 10.86%	-0.0364 10.36%
OBS	608	608	608	608	608	608
R^2	0.6580	0.6564	0.6349	0.6634	0.6713	0.6532

This table presents the summary of regression results for Eqs. (19a), (19b), (20a), and (20b). For the time-series regression models (19a), (19b), (20a), and (20b), the dependent variable is the dividend per share $D_{i,t}$ for firm i at year t . Independent variables are the lag of dividend per share ($D_{i,t-1}$ and $D_{i,t-2}$), current earnings per share ($E_{i,t}$), and permanent earnings per share ($E^P_{i,t}$) for firm i at year t . The independent variable, permanent earnings per share calculated by Darby's method, is used in Eqs. (19b) and (20b). The independent variable, permanent earnings per share calculated by Lee and Primeaux's method, is used in Eqs. (19b)* and (20b)*. Coefficients presented are the cross-sectional averages of estimated coefficients of the time-series regressions. The cross-sectional standard deviations of estimated coefficients of the time-series regressions are in the parenthesis. The medians of estimated coefficients of the time-series regressions are also presented. Percentage numbers show the percentage of significant estimated coefficients of the time-series regressions at 95% significant level. For Eqs. (19a) and (19b), the cross-sectional averages of partial adjustment coefficient and long-term payout are also presented. The cross-sectional averages of the number of observations and R-square for each model are presented at the bottom of table

Table 1B also indicates that the trimmed averages of long-term payout ratios in terms of current EPS and permanent EPS calculated by Darby's method and Lee and Primeaux's method are 0.2600, 0.4603, and 0.3034, respectively. This implies that current EPS instead of permanent EPS is measured with error and estimated regression coefficient is downward biased. It is worthwhile to know that the average short-term payout ratio is 0.3636, which is presented in Appendix C.

Table 1B Partial adjustment coefficient and long-term payout ratios

Variable	Eq. (19a)	Eq. (19b)	Eq. (19b)*
Partial adjustment coefficient	0.4236 (0.2566)	0.5366 (0.2607)	0.4580 (0.2638)
Median	0.3757	0.4943	0.4203
Minimum	0.0015	−0.0004	−0.0977
Maximum	1.2607	1.2965	1.2110
Skewness	0.8207	0.5378	0.5757
Kurtosis	0.2156	−0.3007	−0.3615
Trimmed mean	0.4218	0.5355	0.4571
Long-term payout	0.3115 (0.9007)	−0.3547 (19.9900)	0.4041 2.3635
Median	0.2213	0.3243	0.2151
Minimum	−2.3394	−463.0488	−7.6691
Maximum	15.1902	50.2932	53.4270
Skewness	14.6148	−20.6990	19.1006
Kurtosis	238.3788	476.1219	421.5753
Trimmed mean	0.2660	0.4603	0.3034

This table presents the summary of partial adjustment coefficient and long-term payout ratios for 608 firms. Each firm’s partial adjustment coefficient is equal to one minus the coefficient of the lag of dividend per share ($D_{i,t-1}$) in Eqs. (19a), (19b), and (19b)*. In Eq. (19a), the long-term payout ratio of individual firm is equal to the coefficient of current earnings per share ($E_{i,t}$) divided by its partial adjustment coefficient. In Eqs. (19b) and (19b)*, the long-term payout ratios of individual firm is equal to the coefficient of Darby’s and Lee and Primeaux’s permanent earnings per share ($E^P_{i,t}$) divided by their partial adjustment coefficient, respectively.

The coefficients presented are the cross-sectional averages of partial adjustment coefficients and long-term payout ratios for 608 firms. The cross-sectional standard deviations are in the parenthesis. The median, minimum, and maximum, skewness, kurtosis values of estimated coefficients are also presented. Trimmed mean is calculated by excluding 1% of sample’s extreme value. That is, trimmed mean can be obtained by taking out six outliers of estimated coefficients and then calculating the average of the remaining estimated coefficients

The averages of partial adjustment coefficients in terms of current EPS and permanent EPS calculated by Darby’s method and Lee and Primeaux’s are similar regardless whether regular mean, median, or trimmed mean are used.

Table 2 presents alternative statistical information of current and permanent EPS, payout ratio, and estimated λ_i by using either Darby’s method or Lee and Primeaux’s method. The average EPS from current earnings, permanent earnings by Darby’s method, and permanent earnings by Lee and Primeaux’s method are 2.4290, 2.1867, and 2.3834, respectively. The average payout ratios from current earnings, permanent earnings by Darby’s method, and permanent earnings by Lee and Primeaux’s method are 0.3636, 0.4244, and 0.3722, respectively. The average estimated λ_i by using Darby’s and Lee and Primeaux’s methods are 0.6875 and 0.5795, respectively. This implies that Lee and Primeaux’s method for estimating permanent earnings weights more heavily on current earnings than those from Darby’s method.

70.4.1.2 Results from Pooled Regression

Table 3 presents the results from pooled regression by using both current and permanent EPS calculated by Darby's method and Lee and Primeaux's method. We found that the results from pooled data are similar to the trimmed mean presented in Table 1B. In other words, the estimated intercepts using two alternative permanent EPS measurement are smaller than that of using current EPS and the estimated C_1 in terms of permanent EPS that is larger than that of using current EPS.

70.4.2 Lambrecht and Myer's Method

Since ρ_i is not available for an individual firm, we use a limiting definition of Lambrecht and Myers' (2012) method (see Eq. [15]) to estimate permanent income and apply permanent income to test dividend payment behavior models. More specifically, we estimate the following four dividend payment behavior models:

$$d_{i,t} = a_0 + a_1 TE_{i,t} + a_2 d_{i,t-1} + e_{i,t} \quad (22a)$$

$$d_{i,t} = a_0 + a_1 Y_{i,t} + a_2 d_{i,t-1} + e_{i,t} \quad (22b)$$

$$d_{i,t} = a_0 + a_1 TE_{i,t} + a_2 d_{i,t-1} + a_3 d_{i,t-2} + e_{i,t} \quad (23a)$$

$$d_{i,t} = a_0 + a_1 Y_{i,t} + a_2 d_{i,t-1} + a_3 d_{i,t-2} + e_{i,t} \quad (23b)$$

where $d_{i,t}$ is total dividend payout for firm i at time t , $TE_{i,t}$ is net income for firm i at time t , and $Y_{i,t}$ is permanent income for firm i at time t defined as operating income subtracted by previous year interest expenses.

In addition, Lambrecht and Myers (2012) show that the Lintner model may be subject to the model misspecification. As indicated in Eq. (17), the change of payout can be determined by the net income, the previous dividend payout, the transitory income, and the previous debt outstanding. We therefore test the model misspecification by using Eq. (24):

$$\Delta d_{i,t} = a_0 + a_1 TE_{i,t} + a_2 Y_{i,t} + a_3 d_{i,t-1} + a_4 TD_{i,t-1} + e_{i,t} \quad (24)$$

where $Y_{i,t}$ is the interest expenses for firm i at time t . Empirical results are presented in Table 4 as follows:

Table 4 presents the summary of regression results for models (22a), (22b), (23a), (23b), and (24). Table 4 shows that the estimated regression coefficients associated with current income and permanent income are 62.50% and 75.87% and are significantly different from zero at 5% significant level, respectively. This table also shows that the average R-square of Eq. (22b) is higher than that of Eq. (22a). Similarly, the average R-square of Eq. (23b) is higher than that of Eq. (23a). Such results suggest that permanent earnings introduced by Lambrecht and Myers (2012) do improve the power of dividend behavior models. In addition, we find there are

Table 2 Alternative EPS and payout ratios

	EPS	Payout	EPS	λ_i	Payout	EPS	λ_i	Payout
	Original data		Darby's method		Lee and Primeaux's method		Lee and Primeaux's method	
Mean	2.4290	0.3636	2.1867	0.6875	0.4244	2.3834	0.5795	0.3722
Median	2.2001	0.3644	1.9042	0.85	0.3975	2.1908	0.6243	0.3672
Minimum	-0.0200	-6.5141	0.4005	0	0.0140	-0.0337	0	-3.8642
Maximum	14.3671	1.0493	15.4233	1	1.3211	13.3135	0.9985	1.3380
Variance	1.8977	0.0985	1.7753	0.1142	0.0344	1.6071	0.0618	0.0521
Standard deviation	1.3776	0.3139	1.3324	0.3379	0.1854	1.2677	0.2487	0.2282
Skewness	3.3799	-17.2978	4.1154	-0.8648	1.1500	2.9808	-0.6883	-10.2078
Kurtosis	20.6265	380.6515	29.2884	-0.6692	2.8760	17.1606	-0.1988	196.0971

This table presents statistical analysis of λ_i and permanent EPS calculated by both Darby's and Lee and Primeaux's methods. The payout ratios calculated by current EPS and two alternatives permanent EPS are also presented in Table 2

Table 3 Pooled regression results for Eqs. (19a), (19b), (20a), and (20b)

Dependent variable	Eq. (19a)	Eq. (19b)	Eq. (19b)*	Eq. (20a)	Eq. (20b)	Eq. (20b)*
	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$
Intercept	0.2188** (0.0077)	0.1304** (0.0082)	0.1305** (0.0083)	0.1380** (0.0079)	0.0619** (0.0084)	0.0688** (0.0084)
$E_{i,t}$	0.0928** (0.0020)			0.0857** (0.0020)		
$E^P_{i,t}$		0.1699** (0.0032)	0.1516** (0.0029)		0.1671** (0.0036)	0.1422** (0.0031)
$D_{i,t-1}$	0.5210** (0.0056)	0.4594** (0.0060)	0.4645** (0.0060)	0.3421** (0.0067)	0.3016** (0.0069)	0.3097** (0.0069)
$D_{i,t-2}$				0.2903** (0.0068)	0.2457** (0.0069)	0.2529** (0.0069)
Partial adjustment coefficient	0.4790	0.5406	0.5355			
Long-term payout	0.1936	0.3143	0.2831			
OBS	24,432	24,432	24,432	23,824	23,824	23,824
R^2	0.4453	0.4613	0.4597	0.4926	0.5017	0.4980

This table presents pooled regression results for Eqs. (19a), (19b), (20a), and (20b). For the time-series regression models (19a), (19b), (20a), and (20b), the dependent variable is the dividend per share $D_{i,t}$ for firm i at year t . Independent variables are the lag of dividend per share ($D_{i,t-1}$ and $D_{i,t-2}$), current earnings per share ($E_{i,t}$), and permanent earnings per share ($E^P_{i,t}$) for firm i at year t . This table shows the coefficients and standard errors in the parenthesis. The independent variable, permanent earnings per share calculated by Darby's method, is used in Eqs. (19b) and (20b). The independent variable, permanent earnings per share calculated by Lee and Primeaux's method, is used in Eqs. (19b)* and (20b)*. ** denotes significant estimated coefficients at 99% significant level. In Eqs. (19a) and (19b), the partial adjustment coefficient is equal to one minus the coefficient of the lag of dividend per share ($D_{i,t-1}$). In Eq. (19a), the long-term payout ratio is equal to the coefficient of current earnings per share ($E_{i,t}$) divided by its partial adjustment coefficient. In Eq. (19b), the long-term payout ratio is equal to the coefficient of permanent earnings per share ($E^P_{i,t}$) divided by its partial adjustment coefficient. The numbers of observations and R-square for each model are presented at the bottom of table

25.45% of firms whose dividend payouts can be determined by their interest expenses. It indicates that there exists a specification error in Lintner's model in terms of current earnings.

The empirical results of Table 4 are based upon the measurement of the permanent income, $Y_{i,t}$, that is equal to $K^q\pi_t - \rho TD_{t-1}$. In this measurement, we assume that π_t follows a random walk ($\mu = 1$). However, empirically we find that π_t does not follow random walk and μ is not equal to one. Therefore, our empirical work can only treat qualitative instead of quantitative results. Hence, it is not meaningful to quantitatively calculate the average partial adjustment coefficient and the average long-term payout ratio as we done in Sect. 70.4.1.

Table 4 Individual regression results for Eqs. (22a), (22b), (23a), (23b), and (24)

Dependent variable	Eq. (22a)	Eq. (22b)	Eq. (23a)	Eq. (23b)	Eq. (24)
Intercept	$d_{i,t}$ 6.9795 (47.6555)	$d_{i,t}$ 5.9235 (47.9125)	$d_{i,t}$ 6.9220 (54.1949)	$d_{i,t}$ 5.4458 (49.3600)	$\Delta d_{i,t}$ 5.0467 (40.7184)
Median	0.4616 17.43%	0.1973 14.71%	0.4125 11.84%	0.2154 11.57%	0.3310 15.54%
$TE_{i,t}$	0.0548 (0.1147)		0.0518 (0.1281)		0.0483 (0.1014)
Median	0.0256 62.50%		0.0229 58.39%		0.0248 57.19%
$Y_{i,t}$		0.0489 (0.0743)		0.0473 (0.0881)	
Median		0.0310 75.87%		0.0296 71.40%	
$d_{i,t-1}$	0.8670 (0.2713)	0.8201 (0.2801)	1.0239 (0.5000)	0.9482 (0.5078)	-0.1583 (0.3586)
Median	0.9314 94.90%	0.8787 94.05%	1.0996 91.94%	1.0050 91.07%	-0.0976 39.83%
$d_{i,t-2}$			-0.1598 (0.4672)	-0.1253 (0.5891)	
Median			-0.1902 42.93%	-0.1674 32.73%	
$\rho_i TD_{i,t-1}$					0.0020 (0.7949)
Median					0.0104 25.45%
OBS	608	605	608	605	605
R^2	0.8764	0.8807	0.8837	0.8876	0.4040

This table presents the summary of regression results for five regression models. For the time-series regression models (22a), (22b), (23a), and (23b), the dependent variable is the total dividend payout for firm i at year t . For the time-series regression model (24), the dependent variable is the change of total dividend payout for firm i at year t . Dependent variables are the lag of total dividend payouts ($d_{i,t-1}$ and $d_{i,t-2}$), net income ($TE_{i,t}$), permanent income ($Y_{i,t}$), and total interest payment for firm i at year t . Coefficients presented are the cross-sectional averages of estimated coefficients of the time-series regressions. The cross-sectional standard deviations of estimated coefficients of the time-series regressions are in the parenthesis. Percentage numbers show the percentage of significant estimated coefficients of the time-series regressions at 95% significant level. The cross-sectional averages of the number of observations and R-square for each model are also presented.

70.4.3 Combined Model

70.4.3.1 Results from 605 Individual Regressions

In this section, we will modify Eq. (18) in terms of EPS and DPS as follows:

$$D_{i,t} = b_0 + b_1 E_{i,t} + b_2 D_{i,t-1} + b_3 I_{i,t-1} + u_{i,t} \quad (25a)$$

$$D_{i,t} = b_0 + b_1 E_{i,t}^P + b_2 D_{i,t-1} + b_3 I_{i,t-1} + u_{i,t} \quad (25b)$$

where $D_{i,t}$ and $D_{i,t-1}$ are dividend per share for firm i at time t and $t-1$, respectively; $E_{i,t}$ and $E_{i,t}^P$ are current and permanent EPS for firm i at time t ; and $I_{i,t-1}$ is the interest expense per share firm i at time $t-1$. Please note that Eqs. (25a) and (25b) are similar to Eqs. (19a) and (19b). In other words, we add interest expense per share variable to Eqs. (19a) and (19b) to obtain Eqs. (25a) and (25b). Since there are three firms, Rexam Plc., Telus Corp., and Warwick Valley Telephone Co., which do not have interest expense data, the total sample used in Eqs. (25a) and (25b) contains 605 individual firms.

We also estimate combined model as present in Eqs. (25a) and (25b) in Table 5. The empirical results of Eq. (25a) show that there are 52.23% estimated b_1 , 85.29% estimated b_2 , and 22.64% estimated b_3 significantly different from zero at 5% significant level, respectively. From empirical results of Eq. (25b) by using Darby's method, we found that there are 61.82% estimated b_1 , 69.09% estimated b_2 , and 21.16% estimated b_3 significantly different from zero at 5% significant level, respectively. From empirical results of Eq. (25b) by using Lee and Primeaux's method, we found that there are 49.09% estimated b_1 , 75.70% estimated b_2 , and 22.15% estimated b_3 significantly different from zero at 5% significant level, respectively. In addition, we found that the estimated b_1 from permanent EPS by using both Darby's and Lee and Primeaux's methods is larger than that of current EPS and the estimated intercepts using two alternative permanent EPS measurement are smaller than that of using current EPS. Finally, we found about 22% firms with significant estimated b_3 for both Eqs. (25a) and (25b).

The empirical results presented in Table 5 can be used to test whether the companies' annual EPS is following the random walk or not. In addition, these results might also be used to test whether Lambrecht and Myers's budget constraint is held for individual firm or not. Equation (17) derived by Lambrecht and Myers (2012) is based upon the important budget constraint. Following their paper, we explicitly define the budget constraint as follows:

$$d_t + r_t = \pi_t(K) - \rho TD_{t-1} + (TD_t - TD_{t-1}) \quad (26)$$

where d_t is total dividend payout at time t , TD_t and TD_{t-1} are the total debts in period t and $t-1$, respectively; ρ is interest rate; r_t is managerial rents at time t ; and $\pi_t(K)$ is gross profit at time t .

If debt is kept constant ($\Delta TD = TD_t - TD_{t-1} = 0$), the equilibrium payout and managerial rent policies simply split net income, $\alpha(\pi_t(K) - \rho TD_{t-1})$ to payout and $(1 - \alpha)(\pi_t(K) - \rho TD_{t-1})$ to managerial rents. With these policies, payouts and managerial rents follow net income, always in the ratio $\alpha/(1 - \alpha)$, because all future

Table 5 Individual regression results for Eqs. (25a) and (25b)

Dependent variable	Eq. (25a)	Eq. (25b)	Eq. (25b)*
	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$
Intercept	0.2024 (0.3683)	0.0016 (2.9797)	0.1561 (0.5490)
Median	0.1231 34.05%	0.0577 32.89%	0.1198 32.73%
$E_{i,t}$	0.0771 (0.1114)		
Median	0.0483 52.23%		
$E^P_{i,t}$		0.2699 (2.2572)	0.1257 (0.2782)
Median		0.1692 61.82%	0.0884 49.09%
$D_{i,t-1}$	0.5868 (0.2936)	0.4467 (0.2943)	0.5148 (0.2915)
Median	0.6240 85.29%	0.4723 69.09%	0.5540 75.70%
$I_{i,t-1}$	-0.0256 (1.3754)	0.0102 (1.1712)	0.0008 (1.2813)
Median	-0.0215 22.64%	-0.0189 21.16%	-0.0197 22.15%
OBS	605	605	605
R^2	0.6709	0.6789	0.6617

This table presents the summary of regression results for Eqs. (25a) and (25b). For the time-series regression models (25a) and (25b), the dependent variable is the dividend per share $D_{i,t}$ for firm i at year t . Independent variables are the lag of dividend per share ($D_{i,t-1}$), current earnings per share ($E_{i,t}$), permanent earnings per share ($E^P_{i,t}$), and the lag of interest expense per share ($I_{i,t-1}$) for firm i at year t . The independent variables, permanent earnings per shares calculated by Darby's and Lee and Primeaux's methods, are used in Eqs. (25b) and (25b)*, respectively. Coefficients presented are the cross-sectional averages of estimated coefficients of the time-series regressions. The medians of estimated coefficients of the time-series regressions are also presented. The cross-sectional standard deviations of estimated coefficients of the time-series regressions are in the parenthesis. Percentage numbers show the percentage of significant estimated coefficients of the time-series regressions at 95% significant level. The cross-sectional averages of the number of observations and R-square for each model are presented at the bottom of table

income will also be split in this ratio, outside equity, $S_t = \alpha(V_t(K) - (1 + \rho)TD_{t-1})$, and the present value of managerial rents, $R_t = (1 - \alpha)(V_t(K) - (1 + \rho)TD_{t-1})$. Managers would of course like to reduce payouts and take more rents, but cannot do so without violating the capital market constraint. Managers pay out no more than

necessary, so the capital market constraint pins down payouts, rents, and values exactly.

If the budget constraint does not hold, then the term associated with interest expense will not necessarily exist. Even if the budget constraint holds and the annual EPS follows a random walk, then the interest expense per share item will be dropped out. Our empirical test shows that almost all annual EPS for 605 firms do not follow a random walk. Therefore, the empirical results presented in Table 5 imply that there are only 22.64%, 21.16%, or 22.15% firms where budget constraints hold under the Lambrecht and Myers theoretical model.

Budget constraint presented in Eq. (26) implies that only changes of debt are used to adjust the need of new funds. In other words, there exists no external equity issued for the need of investment expansion for a firm. Higgins (1977, 1981, 2008) have used similar budget constraint to calculate its sustainable growth rate. However, his budget constraint imposes the optimal debt asset ratio. Chen et al. (2013) and Lee et al. (2011) have expanded Higgins' budget constraint by allowing new equity issued as alternative source of funds. Therefore, it may be more realistic to generalize the Eq. (26) in terms of either Higgins' or Chen et al. (2013) budget constraints.

70.4.3.2 Results from Pooled Regression

Using pooled data, we estimate both Eqs. (25a) and (25b) and the empirical results are presented in Table 6. Table 6 shows us that the estimated b_0 and b_1 and b_2 are similar to those estimated without interest expense per share term which can be found in Table 3. However, it is worthwhile to know that the estimated coefficient associated with interest expense per share term is not significantly different from zero at a 5% significant level when the permanent EPS is used. This might imply that the permanent EPS not only can remove random fluctuation of EPS but can also remove parts of misspecification error which is shown by Lambrecht and Myers.

70.5 Summary and Concluding Remarks

Based upon the theories and methods developed by Marsh and Merton (1987), Lee and Primeaux (1991), Garrett and Priestley (2000), and Lambrecht and Myers (2012), in this chapter, we performed both theoretically analyses and empirical studies. We investigated how firms generally allocate permanent earnings and transitory earnings between dividend payments and retained earnings. Building on Friedman's permanent income hypothesis, we first showed how current earnings can be decomposed into permanent and transitory components in terms of methods suggested by Darby (1972, 1974). We then used both Darby's and Lee and Primeaux's methods to decompose current EPS into permanent and transitory components and performed empirical investigations. We found that the average long-term payout ratio is downward biased and the average estimated intercept is upward biased when current instead of permanent EPS are used. In addition, we used Lambrecht and Myers' permanent earnings

Table 6 Pooled regression results for Eqs. (25a) and (25b)

Dependent variable	Eq. (25a)	Eq. (25b)	Eq. (25b)*
	$D_{i,t}$	$D_{i,t}$	$D_{i,t}$
Intercept	0.2009** (0.0082)	0.1363** (0.0085)	0.1329** (0.0084)
$E_{i,t}$	0.0920** (0.0020)		
$E^P_{i,t}$		0.1713** (0.0033)	0.1526** (0.0030)
$D_{i,t-1}$	0.5083** (0.0059)	0.4484** (0.0001)	0.4586** (0.0061)
$I_{i,t-1}$	0.0316** (0.0033)	-2.6E-06 (3.7E-06)	-1.6E-06 (3.6E-06)
OBS	24,299	24,299	24,299
R^2	0.4431	0.4445	0.4564

This table presents pooled regression results for Eqs. (25a) and (25b). For the time-series regression models (25a) and (25b), the dependent variable is the dividend per share $D_{i,t}$ for firm i at year t . Independent variables are the lag of dividend per share ($D_{i,t-1}$), current earnings per share ($E_{i,t}$), permanent earnings per share ($E^P_{i,t}$), and the lag of interest expense per share ($I_{i,t-1}$) for firm i at year t . The independent variables, permanent earnings per shares calculated by Darby's and Lee and Primeaux's methods, are used in Eqs. (25b) and (25b)*, respectively. This table shows the coefficients and standard errors in the parenthesis. ** denotes significant estimated coefficients at 99% significant level. The numbers of observations and R-square for each model are presented at the bottom of table

measurement to estimate dividend behavior model. We found that their permanent earnings measurement performs better than the current earnings measurement. However, the permanent earnings measurements from Lambrecht and Myers' method are difficult to be empirically measured in terms of accounting data. Finally, we also empirically investigated the misspecification issue presented by Lambrecht and Myers and found that interest expense per share might be useful for estimating dividend behavior model for some firms.

Based upon the partial adjusted model and the adaptive expectation model, and the integration of these models, we theoretically developed and empirically investigated both current and permanent dividend payout behavioral models. We analyzed these two dividend behavior models by data of individual firms and pooled data. Empirical results show that it is better to use permanent EPS instead of current EPS to estimate dividend behavioral models. If we use current EPS instead of permanent EPS, the estimated intercept will be upward biased and the long-term payout ratio will be underestimated.

In future research, we will first revise the permanent earnings measurement developed by Lambrecht and Myers to make it more plausible for using accounting data to conduct empirical studies for examining dividend behavior. Secondly, we will extend Marsh and Merton's (1987) and Garrett and Priestley's (2000) theories

and models from aggregate dividend behavior models to individual dividend behavior models to test either the signaling theory hypothesis or the free cash flow hypothesis for individual firms.

70.6 Cross-References

- [Optimal Payout Ratio Under Uncertainty and the Flexibility Hypothesis: Theory and Empirical Evidence](#)

Appendix A: Detailed Definition of Permanent Income

In Eq. (11) of Lambrecht and Myers (2012), they define permanent income as:

$$Y_t = \rho\beta \sum_{j=0}^{\infty} \beta^j \sum_{j=0}^{\infty} \beta^j E_t [K^\phi \pi_{t+j}(\eta_{t+j})] \rho D_{t-1} \quad (\text{A1})$$

where ρ and β are interest rate and discount factor, respectively; $E_t[\cdot]$ is the expectation operator; and $K^\phi \pi_{t+j}$ is total net income without corporate tax in period $t+j$.

π_t is gross profit at time t that follows the AR(1) process $\pi_t = \mu\pi_{t-1} + \eta_t$ with $\mu \in [0, 1]$. The shocks η_{t+j} ($j = 0, 1, \dots$) are independently and identically normally distributed with zero mean and volatility σ_η .

Permanent income Y_t defined in Eq. (A1) is the rate of return on the sum of current income and the present value of all future income, net of debt service, but before rents. It is an annuity payment that, given expectations at time t , could be sustained forever. By using AR(1) process discussed in previous paragraph, Lambrecht and Myers claim that Eq. (15) can be derived from Eq. (A1).

This permanent income measurement defined in Eq. (A1) does not take into account a corporate tax. In addition, administration and sales expense were not explicitly considered. Since the budget constraint used to derive this permanent income measurement does not allow new equity, therefore this kind of permanent income measurement has some limitations. In sum, the permanent income measure defined in either Eq. (15) or Eq. (A1) is not exactly followed the permanent income concept developed by Friedman (1957), Darby (1972, 1974), and Wang (2003).

Appendix B: Impacts of Measurement Errors on Estimated Regression Coefficients

By using Lee et al. (2015) and Lee et al. (2019) notations and specification equations, suppose we have a trivariate structural relationship

$$W_i = \alpha + \beta U_i + \gamma V_i \quad (\text{B1})$$

W_i , U_i , and V_i are unobserved, but we can observe $Z_i = W_i + \tau_i$, $X_i = U_i + \varepsilon_i$, and $Y_i = V_i + \eta_i$. U_i and V_i have a joint normal distribution with variances σ_U^2 and σ_V^2 and correlation coefficient ρ_{UV} . In the observed variables X , Y , and Z , the observed errors ε , η , and τ are independent normal variables with zero means and variance σ_1^2 , σ_2^2 , σ_3^2 . X , Y , and Z have a multivariate normal distribution with parameters as follows:

$$\begin{aligned}
 & \text{(a) } m_1 = E(X) \\
 & \text{(b) } m_2 = E(Y) \\
 & \text{(c) } m_3 = \alpha + \beta m_1 + \gamma m_2 \\
 & \text{(d) } m_{XX} = \text{Var}(X) = \sigma_U^2 + \sigma_1^2 \\
 & \text{(e) } m_{YY} = \text{Var}(Y) = \sigma_V^2 + \sigma_2^2 \\
 & \text{(f) } m_{ZZ} = \text{Var}(Z) = \beta^2 \sigma_U^2 + \gamma^2 \sigma_V^2 + 2\beta\gamma\rho_{UV}\sigma_U\sigma_V + \sigma_3^2 \\
 & \text{(g) } m_{XY} = \rho_{UV}\sigma_U\sigma_V \\
 & \text{(h) } m_{XZ} = \beta\sigma_U^2 + \gamma\rho_{UV}\sigma_U\sigma_V \\
 & \text{(i) } m_{YZ} = \beta\rho_{UV}\sigma_U\sigma_V + \gamma\sigma_V^2.
 \end{aligned} \tag{B2}$$

The joint sufficient statistics of m_1 , m_2 , m_3 , m_{XX} , m_{YY} , m_{ZZ} , m_{XY} , m_{XZ} , and m_{YZ} can be defined as

$$\begin{aligned}
 & \text{(a) } \bar{X} = \frac{\sum_{i=1}^n X_i}{n} \\
 & \text{(b) } \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \\
 & \text{(c) } \bar{Z} = \frac{\sum_{i=1}^n Z_i}{n} \\
 & \text{(d) } S_{XX} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \\
 & \text{(e) } S_{YY} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n} \\
 & \text{(f) } S_{ZZ} = \frac{\sum_{i=1}^n (Z_i - \bar{Z})^2}{n} \\
 & \text{(g) } S_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n} \\
 & \text{(h) } S_{XZ} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Z_i - \bar{Z})}{n} \\
 & \text{(i) } S_{YZ} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})}{n}
 \end{aligned} \tag{B3}$$

From Eqs. (B2) and (B3), we know that $\bar{X}$, $\bar{Y}$, $\bar{Z}$, S_{XX} , S_{YY} , S_{ZZ} , S_{XY} , S_{XZ} , and S_{YZ} are joint sufficient statistics of m_1 , m_2 , m_3 , m_{XX} , m_{YY} , m_{ZZ} , m_{XY} , m_{XZ} , and m_{YZ} . If the

former nine variables are jointly independent a set of maximum likelihood equations can be formulated as follows:

$$\begin{aligned}
 (a) \quad S_{XX} &= \sigma_U^2 + \sigma_1^2 \\
 (b) \quad S_{YY} &= \sigma_V^2 + \sigma_2^2 \\
 (c) \quad S_{ZZ} &= \hat{\beta}^2 \sigma_U^2 + \hat{\gamma}^2 \sigma_V^2 + 2\hat{\beta}\hat{\gamma}\sigma_{UV} + \sigma_3^2 \\
 (d) \quad S_{XY} &= \sigma_{UV} \\
 (e) \quad S_{XZ} &= \hat{\beta}\sigma_U^2 + \hat{\gamma}\sigma_{UV} \\
 (f) \quad S_{YZ} &= \hat{\beta}\sigma_{UV} + \hat{\gamma}\sigma_V^2
 \end{aligned} \tag{B4}$$

Equations (B1), (B2), (B3), and (B4) will be used in determining effects of measurement errors on regression coefficients.

From Eq. (B4), the effects of measurement errors on the estimates of β and γ can be seen from the following:

$$\text{plim } \hat{\beta} = \frac{(\sigma_V^2 + \sigma_2^2)\sigma_{WV} - (\sigma_{WV}\sigma_{UV})}{(\sigma_U^2 + \sigma_1^2)(\sigma_V^2 + \sigma_2^2) - (\sigma_{WV})^2} \tag{B5}$$

$$\text{plim } \hat{\gamma} = \frac{(\sigma_V^2 + \sigma_1^2)\sigma_{WV} - (\sigma_{WV}\sigma_{UV})}{(\sigma_U^2 + \sigma_1^2)(\sigma_V^2 + \sigma_2^2) - (\sigma_{UV})^2} \tag{B6}$$

From both (B5) and (B6), the asymptotic biases of $\hat{\beta}$ and $\hat{\gamma}$ can be defined as:

$$\text{plim } \hat{\beta} - \beta = \frac{\sigma_{VW}\sigma_2^2 - \beta(\sigma_U^2\sigma_2^2 + \sigma_V^2\sigma_1^2 + \sigma_1^2\sigma_2^2)}{(\sigma_U^2\sigma_V^2 - \sigma_{VW}^2) + \sigma_U^2\sigma_2^2 + \sigma_V^2\sigma_1^2 + \sigma_1^2\sigma_2^2} \tag{B7}$$

$$\text{plim } \hat{\gamma} - \gamma = \frac{\sigma_{WV}\sigma_1^2 - \gamma(\sigma_U^2\sigma_2^2 + \sigma_V^2\sigma_1^2 + \sigma_1^2\sigma_2^2)}{(\sigma_U^2\sigma_V^2 - \sigma_{UV}^2) + \sigma_U^2\sigma_2^2 + \sigma_V^2\sigma_1^2 + \sigma_1^2\sigma_2^2} \tag{B8}$$

The direction of the biases of $\hat{\beta}$ and $\hat{\gamma}$ can be treated according to the following:

- Under the assumption that $\text{Cov}(UV) = 0$

- (i) If $\sigma_1^2 = 0$, $\sigma_2^2 > 0$,

$$\begin{aligned}
 (a) \quad \text{plim } \hat{\beta} - \beta &= \frac{\sigma_2^2(\sigma_{WU} - \beta\sigma_U^2)}{\sigma_U^2(\sigma_V^2 + \sigma_2^2)} = 0, \\
 (b) \quad \text{plim } \hat{\gamma} - \gamma &= \frac{-\gamma\sigma_U^2\sigma_2^2}{\sigma_U^2(\sigma_V^2 + \sigma_2^2)} = \frac{-\gamma\sigma_2^2}{(\sigma_V^2 + \sigma_2^2)}.
 \end{aligned} \tag{B9}$$

Eq. (B9a) implies that $\hat{\beta}$ is an asymptotic unbiased estimator of β , while Eq. (B9b) implies that $\hat{\gamma}$ is downward biased estimator of γ .

(ii) If $\sigma_1^2 > 0$, $\sigma_2^2 = 0$,

$$\begin{aligned} \text{(a) } \text{plim } \hat{\beta} - \beta &= \frac{-\beta\sigma_1^2}{(\sigma_U^2 + \sigma_1^2)}, \\ \text{(b) } \text{plim } \hat{\gamma} - \gamma &= \frac{\sigma_1^2(\sigma_{WV} - \gamma\sigma_V^2)}{\sigma_V^2(\sigma_U^2 + \sigma_1^2)} = 0. \end{aligned} \quad (\text{B10})$$

In accordance with Eq. (B9), Eq. (B10) can be used to draw some meaningful conclusions about the biases of both $\hat{\beta}$ and $\hat{\gamma}$.

(iii) Finally, if $\sigma_1^2 > 0$, $\sigma_2^2 > 0$,

$$\begin{aligned} \text{(a) } \text{plim } \hat{\beta} - \beta &= -\frac{\beta\sigma_1^2}{\sigma_U^2 + \sigma_1^2}, \\ \text{(b) } \text{plim } \hat{\gamma} - \gamma &= -\frac{\gamma\sigma_2^2}{\sigma_V^2 + \sigma_2^2}. \end{aligned} \quad (\text{B11})$$

In this case, both $\hat{\beta}$ and $\hat{\gamma}$ are downward biased estimators of β and γ .

- Suppose that $\text{Cov}(UV) \neq 0$

(i) If $\sigma_1^2 = 0$, $\sigma_2^2 > 0$,

$$\begin{aligned} \text{(a) } \text{plim } \hat{\beta} - \beta &= \gamma b_{VU} \left(\frac{\sigma_2^2}{\sigma_2^2 + \sigma_V^2(1 - R_{UV}^2)} \right), \\ \text{(b) } \text{plim } \hat{\gamma} - \gamma &= \frac{-\gamma\sigma_2^2}{\sigma_V^2 - b_{VU} + \sigma_2^2}, \end{aligned} \quad (\text{B12})$$

where b_{VU} is the auxiliary regression coefficient of a regressing V on U , and R_{UV}^2 is the correlation coefficient between U and V .

(B12a) implies that the direction of the bias of $\hat{\beta}$ depends upon the sign of both γ and b_{VU} ; (B12b) implies that γ is a downward biased estimator of γ unless $(\sigma_V^2 - b_{VU} + \sigma_2^2)$ is smaller than zero.

(ii) If $\sigma_1^2 > 0$, $\sigma_2^2 = 0$,

$$\begin{aligned}
 \text{(a) } \text{plim } \hat{\beta} - \beta &= \frac{-\beta\sigma_1^2}{\sigma_U^2 - b_{UV} + \sigma_1^2}, \\
 \text{(b) } \text{plim } \hat{\gamma} - \gamma &= \beta b_{UV} \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_U^2(1 - R_{UV}^2)} \right),
 \end{aligned}
 \tag{B13}$$

where b_{UV} = the auxiliary regression coefficient of regressing U on V .

(iii) If $\sigma_1^2 > 0$, $\sigma_2^2 > 0$,

$$\begin{aligned}
 \text{(a) } \text{plim } \hat{\beta} - \beta &= \frac{\gamma b_{VU} - \frac{\beta}{\sigma_U^2} (\sigma_V^2 \sigma_1^2 + \sigma_1^2 \sigma_2^2)}{\sigma_V^2 - b_{UV} + \sigma_2^2 + \frac{(\sigma_V^2 \sigma_1^2 + \sigma_1^2 \sigma_2^2)}{\sigma_U^2}}, \\
 \text{(b) } \text{plim } \hat{\gamma} - \gamma &= \frac{\beta b_{UV} - \frac{\gamma}{\sigma_V^2} (\sigma_U^2 \sigma_2^2 + \sigma_1^2 \sigma_2^2)}{\sigma_U^2 - b_{UV} + \sigma_1^2 + \frac{\sigma_V^2 \sigma_1^2 + \sigma_1^2 \sigma_2^2}{\sigma_V^2}}.
 \end{aligned}
 \tag{B14}$$

From (B14), we can see that the direction of the biases of both $\hat{\beta}$ and $\hat{\gamma}$ are ambiguous.

Appendix C: EPS, DPS, and Payout Ratio for 608 Firms

Company	EPS	DPS	Payout ratio
WEYERHAEUSER CO	1.8281	1.9182	1.0493
REXAM PLC	0.7033	0.7177	1.0205
GENERAL MOTORS CO	3.2904	3.0041	0.9130
PENNZENERGY CO	2.1978	1.9520	0.8882
ENBRIDGE INC	2.6175	2.2439	0.8573
IMPERIAL CHEMICAL INDUSTRIES PLC	2.0892	1.7763	0.8502
NEWMONT MINING CORP	2.0498	1.5844	0.7730
POTLATCH CORP NEW	2.4045	1.8104	0.7529
THOMSON REUTERS CORP	0.8430	0.6328	0.7507
G T E CORP	2.8789	2.1105	0.7331
TASTY BAKING CO	0.7909	0.5737	0.7254
SOUTHERN NEW ENGLAND TELECOM	3.7937	2.6804	0.7065
WD 40 CO	1.8813	1.3226	0.7030
CHICAGO RIVET & MACH CO	1.8433	1.2807	0.6948
SPRINT NEXTEL CORP	1.6162	1.1128	0.6885

(continued)

Company	EPS	DPS	Payout ratio
A T & T CORP	3.3605	2.2991	0.6842
TEXACO INC	3.9164	2.6687	0.6814
GRACE W R & CO DEL NEW	2.6294	1.7827	0.6780
SERVIDYNE INC	0.2868	0.1931	0.6732
SNYDERS LANCE INC	1.3684	0.9138	0.6678
EASTMAN KODAK CO	2.9343	1.9552	0.6663
B P PLC	3.2504	2.1088	0.6488
AVON PRODUCTS INC	2.3906	1.5341	0.6417
ALBERTO CULVER CO NEW	1.4772	0.9351	0.6330
SKYLINE CORP	0.9220	0.5832	0.6325
CHEMED CORP NEW	1.8529	1.1710	0.6320
TAMBRANDS INC	3.3763	2.1217	0.6284
LOUISIANA LAND & EXPLORATION CO	1.7116	1.0755	0.6284
TRUE NORTH COMMUNICATIONS INC	1.9754	1.2317	0.6235
THOMAS & BETTS CORP	2.3713	1.4757	0.6223
FRONTIER CORP	2.3231	1.4367	0.6184
MAYTAG CORP	2.0576	1.2677	0.6161
U S T INC	2.5000	1.5349	0.6139
WARNER LAMBERT CO	2.4624	1.4859	0.6035
JOSLYN CORP	2.1130	1.2741	0.6030
LAWTER INTERNATIONAL INC	0.7859	0.4714	0.5998
P M F G INC	0.8877	0.5314	0.5986
DUN & BRADSTREET CORP DEL NEW	2.7844	1.6614	0.5967
FLEMING COMPANIES INC	1.3131	0.7779	0.5924
WYETH	3.0332	1.7929	0.5911
RIO ALGOM MINES LTD	1.7808	1.0483	0.5887
ALLTEL CORP	2.2489	1.3191	0.5866
FINA INC	4.4227	2.5854	0.5846
BRISTOL MYERS SQUIBB CO	2.7614	1.6115	0.5836
QUAKER STATE CORP	1.2058	0.7024	0.5825
INCO LTD	1.9033	1.1069	0.5815
GENERAL DYNAMICS CORP	4.7283	2.7483	0.5812
FEDERAL MOGUL CORP	1.8853	1.0931	0.5798
WINN DIXIE STORES INC	2.4259	1.4047	0.5790
GLAXOSMITHKLINE PLC	1.4618	0.8453	0.5783
DOW JONES & CO INC	1.6414	0.9473	0.5772
MOORE WALLACE INC	1.8051	1.0398	0.5760
CURTISS WRIGHT CORP	2.7887	1.6032	0.5749
G A T X CORP	2.6064	1.4951	0.5736
MONEYGRAM INTERNATIONAL INC	1.7273	0.9819	0.5685
ATLANTIC RICHFIELD CO	5.7600	3.2516	0.5645
COCA COLA BOTTLING CO CONS	1.5403	0.8626	0.5601
COURTAULDS PLC	0.3780	0.2104	0.5566
GOODRICH CORP	2.4206	1.3324	0.5504

(continued)

Company	EPS	DPS	Payout ratio
B C E INC	3.4605	1.8965	0.5480
RAYONIER INC NEW	2.5671	1.4014	0.5459
OLIN CORP	2.2860	1.2451	0.5447
C C H INC	1.6973	0.9234	0.5441
LILLY ELI & CO	3.0884	1.6768	0.5429
NATIONAL PRESTO INDS INC	4.0609	2.2036	0.5426
AMERICAN CYANAMID CO	2.8244	1.5315	0.5423
STURM RUGER & CO INC	2.2617	1.2239	0.5412
N L INDUSTRIES INC	1.6049	0.8682	0.5410
NALCO CHEMICAL CO	1.9924	1.0713	0.5377
TIME WARNER INC	1.9360	1.0365	0.5354
NORTH PITTSBURGH SYSTEMS INC	2.3774	1.2637	0.5316
XEROX CORP	3.6154	1.9065	0.5273
WACKENHUT CORP	1.0106	0.5316	0.5260
KELLOGG CO	2.3492	1.2353	0.5259
3 M CO	4.2612	2.2393	0.5255
OUTBOARD MARINE CORP	1.5629	0.8209	0.5253
BEAM INC	4.1410	2.1662	0.5231
INTERNATIONAL PAPER CO	2.8700	1.5012	0.5231
INTERNATIONAL FLAVORS & FRAG INC	2.1183	1.0955	0.5171
GUARDSMAN PRODUCTS INC	0.8579	0.4429	0.5162
NOVA CHEMICALS CORP	0.8083	0.4170	0.5159
ALIAN COMMUNICATIONS INC	2.6455	1.3632	0.5153
BASSETT FURNITURE INDUSTRIES INC	1.9407	0.9992	0.5149
OFFICEMAX INC NEW	1.9137	0.9820	0.5132
ARMSTRONG HOLDINGS INC	2.2338	1.1409	0.5108
COCA COLA CO	2.8166	1.4373	0.5103
MCGRAW HILL COS INC	2.5002	1.2744	0.5097
MOBIL CORP	5.6085	2.8482	0.5078
DU PONT E I DE NEMOURS & CO	5.4174	2.7450	0.5067
MUELLER PAUL CO	3.1856	1.6141	0.5067
STANDARD REGISTER CO	1.8577	0.9410	0.5066
UNION CAMP CORP	3.5191	1.7794	0.5056
GOLDEN ENTERPRISES INC	0.6270	0.3167	0.5050
BETZDEARBORN INC	1.9506	0.9830	0.5039
FREEPORT MCMORAN INC	2.1176	1.0661	0.5035
GILLETTE CO	2.4198	1.2132	0.5014
BLOCK H & R INC	1.9894	0.9955	0.5004
OCCIDENTAL PETROLEUM CORP	2.9730	1.4862	0.4999
BESTFOODS	3.7511	1.8732	0.4994
CLOROX CO	2.3353	1.1620	0.4976
RHONE POULENC RORER INC	1.8354	0.9126	0.4972
HERCULES INC	2.7869	1.3825	0.4961
LUFKIN INDUSTRIES INC	6.3734	3.1593	0.4957

(continued)

Company	EPS	DPS	Payout ratio
KONINKLIJKE PHILIPS ELEC N V	1.8128	0.8982	0.4955
WITCO CORP	2.6251	1.2967	0.4939
TIMKEN COMPANY	2.7931	1.3749	0.4922
GENUINE PARTS CO	2.2388	1.1019	0.4922
TELUS CORP	2.7542	1.3542	0.4917
LINCOLN ELECTRIC HOLDINGS INC	9.3203	4.5808	0.4915
GENERAL MILLS INC	2.9958	1.4666	0.4895
DELUXE CORP	2.3064	1.1266	0.4885
MERCK & CO INC NEW	3.4755	1.6972	0.4883
COVANTA ENERGY CORP	2.5877	1.2605	0.4871
GERBER PRODUCTS CO	2.5927	1.2617	0.4866
MCDERMOTT INTERNATIONAL INC	2.2614	1.1000	0.4864
GANNETT INC	2.1633	1.0512	0.4859
REYNOLDS METALS CO	2.6443	1.2836	0.4854
PHARMACIA CORP	4.5510	2.2065	0.4848
LONE STAR INDUSTRIES INC	2.6590	1.2869	0.4840
WELLCO ENTERPRISES INC	1.0457	0.5054	0.4833
SUREWEST COMMUNICATIONS	1.4568	0.7031	0.4826
COLGATE PALMOLIVE CO	2.5366	1.2230	0.4821
SAMES CORP	1.9100	0.9202	0.4818
DOW CHEMICAL CO	3.7966	1.8224	0.4800
CINCINNATI BELL INC NEW	3.1581	1.5121	0.4788
ESPEY MANUFACTURING & ELCTRS COR	1.4621	0.6992	0.4782
HICKORY TECH CORP	3.0095	1.4376	0.4777
BOWL AMERICA INC	0.9282	0.4432	0.4775
PHELPS DODGE CORP	5.4278	2.5872	0.4767
AMOCO CORP	5.3452	2.5470	0.4765
PFIZER INC	2.3843	1.1320	0.4748
DONNELLEY R R & SONS CO	1.8441	0.8741	0.4740
UNILEVER PLC	3.5507	1.6827	0.4739
EMERSON ELECTRIC CO	3.0286	1.4219	0.4695
T R W INC	3.7608	1.7632	0.4688
BRIGGS & STRATTON CORP	2.7541	1.2888	0.4680
NATIONAL SERVICE INDUSTRIES INC	2.0316	0.9502	0.4677
CHESAPEAKE CORP VA	2.1479	1.0016	0.4663
PHARMACIA & UPJOHN INC	2.9657	1.3780	0.4646
PROCTER & GAMBLE CO	4.0902	1.9005	0.4646
CARPENTER TECHNOLOGY CORP	3.2515	1.5100	0.4644
KERR MCGEE CORP	2.8410	1.3182	0.4640
BROWN SHOE CO INC NEW	2.3751	1.1011	0.4636
AEROQUIP VICKERS INC	2.7791	1.2840	0.4620
KROGER COMPANY	2.6903	1.2428	0.4620
KNAPPE & VOGT MANUFACTURING CO	1.7766	0.8199	0.4615
INGERSOLL RAND PLC	3.2769	1.5103	0.4609

(continued)

Company	EPS	DPS	Payout ratio
UNILEVER N V	6.8549	3.1517	0.4598
HUBBELL INC	2.8068	1.2875	0.4587
VULCAN INTERNATIONAL CORP	1.2687	0.5798	0.4570
AVERY DENNISON CORP	1.7728	0.8099	0.4568
SALIENT 3 COMMUNICATIONS INC	2.2555	1.0303	0.4568
CAMPBELL SOUP CO	2.5189	1.1500	0.4565
DOMINION TEXTILE LTD	1.2726	0.5800	0.4558
WRIGLEY WILLIAM JR CO	4.4873	2.0438	0.4555
LONGVIEW FIBRE CO	5.9690	2.7092	0.4539
EXXON MOBIL CORP	5.7540	2.6115	0.4538
FRIEDMAN INDUSTRIES INC	0.8190	0.3716	0.4537
HERSHEY CO	2.4530	1.1114	0.4531
BLAIR CORP	2.2326	1.0111	0.4529
PITNEY BOWES INC	2.4168	1.0941	0.4527
INTERNATIONAL MULTIFOODS CORP	2.4650	1.1144	0.4521
WOODHEAD INDUSTRIES INC	0.9318	0.4210	0.4518
ALTRIA GROUP INC	4.7865	2.1537	0.4499
HOMESTAKE MINING CO	1.2123	0.5452	0.4497
HEINZ H J CO	2.9502	1.3259	0.4494
KIMBERLY CLARK CORP	4.4500	1.9977	0.4489
SUNDSTRAND CORP	2.7809	1.2443	0.4474
ABITIBI CONSOLIDATED INC	1.2510	0.5595	0.4472
PENNEY J C CO INC	3.2683	1.4613	0.4471
MCKESSON H B O C INC	2.3373	1.0448	0.4470
SEARS ROEBUCK & CO	3.1981	1.4277	0.4464
CHEVRON CORP NEW	5.5898	2.4954	0.4464
GEORGIA PACIFIC CORP	2.3538	1.0504	0.4462
FOOT LOCKER INC	2.5685	1.1441	0.4454
MEADWESTVACO CORP	2.3265	1.0362	0.4454
HANNA M A CO DE	2.1487	0.9561	0.4450
SUNOCO INC	3.4877	1.5499	0.4444
ENESCO GROUP INC	2.6219	1.1643	0.4441
GENERAL ELECTRIC CO	3.5628	1.5821	0.4441
GORMAN RUPP CO	1.7089	0.7587	0.4440
HOONEYWELL INTERNATIONAL INC	3.0808	1.3611	0.4418
ROYAL DUTCH PETROLEUM CO	8.4524	3.7332	0.4417
GARAN INC	2.3678	1.0455	0.4416
SPRINGS INDUSTRIES INC	2.5746	1.1354	0.4410
R P M INTERNATIONAL INC	1.0638	0.4689	0.4408
MILACRON INC	1.7380	0.7651	0.4402
A M P INC	2.1559	0.9462	0.4389
CHEMTURA CORP	1.3735	0.6023	0.4385
PRATT & LAMBERT UNITED INC	1.8613	0.8160	0.4384
LUKENS INC DE	2.1091	0.9220	0.4371

(continued)

Company	EPS	DPS	Payout ratio
P P G INDUSTRIES INC	3.9091	1.7072	0.4367
POPE & TALBOT INC	1.5326	0.6664	0.4348
MARCUS CORP	1.1427	0.4955	0.4336
ARKANSAS BEST CORP DEL	0.9140	0.3953	0.4325
MALLINCKRODT INC NEW	3.2615	1.4060	0.4311
SUPERVALU INC	1.4922	0.6431	0.4310
C B S CORP	2.5566	1.1002	0.4303
S P X CORP	2.8147	1.2090	0.4295
CONSOLIDATED PAPERS INC	3.7724	1.6197	0.4293
VULCAN MATERIALS CO	3.7418	1.6059	0.4292
FEDERAL SIGNAL CORP	1.3832	0.5935	0.4291
GOODYEAR TIRE & RUBBER CO	2.7623	1.1849	0.4290
CHURCHILL DOWNS INC	3.5490	1.5186	0.4279
ANGELICA CORP	1.2422	0.5310	0.4275
ECHLIN INC	1.3366	0.5700	0.4265
SENSIENT TECHNOLOGIES CORP	1.9745	0.8390	0.4249
HARSCO CORP	2.6972	1.1416	0.4233
INTERNATIONAL BUSINESS MACHS COR	7.9610	3.3628	0.4224
TIMES MIRROR CO NEW	2.3384	0.9864	0.4218
DEXTER CORP	1.7712	0.7459	0.4211
ERICSSON L M TELEPHONE CO	2.1149	0.8889	0.4203
STANLEY BLACK & DECKER INC	2.4594	1.0311	0.4192
BRUNSWICK CORP	1.1472	0.4807	0.4190
BARNES GROUP INC	2.2872	0.9577	0.4187
KEWAUNEE SCIENTIFIC CORP	0.9451	0.3932	0.4160
SCHERING PLOUGH CORP	2.5168	1.0448	0.4151
NORFOLK SOUTHERN CORP	4.8251	2.0003	0.4146
INTERNATIONAL ALUMINUM CORP	1.8497	0.7665	0.4144
SNAP ON INC	2.2218	0.9198	0.4140
BLESSINGS CORP	1.4512	0.6007	0.4139
BROWN FORMAN CORP	2.8786	1.1911	0.4138
DANA HOLDING CORP	2.8242	1.1661	0.4129
TWIN DISC INC	1.8986	0.7830	0.4124
ALCAN INC	2.1169	0.8687	0.4104
NEWELL RUBBERMAID INC	1.5415	0.6325	0.4103
I T T CORP	3.1258	1.2796	0.4094
GENCORP INC	2.2339	0.9082	0.4065
KUBOTA CORP	1.6162	0.6562	0.4060
PENN VIRGINIA CORP	3.0980	1.2516	0.4040
GOULDS PUMPS INC	2.0306	0.8196	0.4036
MARSH SUPERMARKETS INC	1.0505	0.4227	0.4024
LOUISIANA PACIFIC CORP	1.5763	0.6336	0.4020
CONAGRA INC	1.8431	0.7399	0.4015
HANDLEMAN CO	1.4835	0.5935	0.4000

(continued)

Company	EPS	DPS	Payout ratio
CUMMINS INC	3.2872	1.3130	0.3994
ABBOTT LABORATORIES	2.5606	1.0181	0.3976
CATERPILLAR INC	3.3917	1.3479	0.3974
SMUCKER J M CO	2.3951	0.9516	0.3973
CALIBER SYSTEM INC	2.2680	0.8990	0.3964
GENESIS WORLDWIDE INC	1.7586	0.6953	0.3954
SAVANNAH FOODS & INDUSTRIES INC	2.9056	1.1462	0.3945
UNITED STATES SHOE CORP	2.1310	0.8402	0.3943
TENNANT CO	2.0393	0.7993	0.3920
POLAROID CORP	1.7000	0.6653	0.3914
HILTON HOTELS CORP	2.3536	0.9201	0.3909
GRUMMAN CORP	2.3890	0.9337	0.3908
CASTLE A M & CO	1.8077	0.7057	0.3904
MACYS INC	3.2067	1.2512	0.3902
NASH FINCH COMPANY	2.1303	0.8297	0.3895
EASTERN CO	1.8702	0.7284	0.3895
WEIS MARKETS INC	2.2283	0.8675	0.3893
FLOWERS FOODS INC	1.3039	0.5072	0.3890
MAY DEPARTMENT STORES CO	3.3091	1.2868	0.3889
QUAKER OATS CO	3.1665	1.2296	0.3883
FOSTER WHEELER AG	1.6951	0.6582	0.3883
BAXTER INTERNATIONAL INC	1.5902	0.6169	0.3880
FERRO CORP	2.0320	0.7877	0.3876
UNIVERSAL CORPORATION	3.4365	1.3259	0.3858
DIEBOLD INC	2.2038	0.8493	0.3854
PEPSIAMERICAS INC NEW	2.2533	0.8665	0.3846
TEXTRON INC	3.1694	1.2187	0.3845
SCOTT PAPER CO	2.2688	0.8715	0.3841
CONOCOPHILLIPS	3.7772	1.4493	0.3837
ROLLINS INC	0.9864	0.3784	0.3837
ARVIN INDUSTRIES INC	1.9923	0.7634	0.3832
PEPSICO INC	2.6327	1.0072	0.3826
ENCANA CORP	2.4120	0.9214	0.3820
HANDY & HARMAN	1.4874	0.5677	0.3817
ROCKWELL AUTOMATION INC	3.1443	1.1982	0.3811
MARION MERRELL DOW INC	1.2747	0.4855	0.3809
LINDBERG CORP	0.9938	0.3784	0.3808
ROANOKE ELECTRIC STEEL CORP	1.9797	0.7538	0.3808
BAUSCH & LOMB INC	2.3069	0.8772	0.3803
TECUMSEH PRODUCTS CO	7.2580	2.7551	0.3796
VELCRO INDUSTRIES N V	2.2407	0.8488	0.3788
AMCAST INDUSTRIAL CORP	2.2197	0.8343	0.3759
OXFORD INDUSTRIES INC	1.6317	0.6133	0.3759
BADGER METER INC	1.5669	0.5876	0.3750

(continued)

Company	EPS	DPS	Payout ratio
SEARS HOLDINGS CORP	2.3200	0.8695	0.3748
STRIDE RITE CORP	1.5433	0.5782	0.3746
UNITED TECHNOLOGIES CORP	4.2492	1.5892	0.3740
KYSOR INDUSTRIAL CORP DE	1.5783	0.5901	0.3739
UNITED STATES STEEL CORP NEW	3.4600	1.2910	0.3731
COOPER INDUSTRIES PLC	3.4150	1.2730	0.3728
STONE & WEBSTER INC	3.6697	1.3670	0.3725
BEMIS CO INC	2.4298	0.9046	0.3723
DI GIORGIO CORP	1.2227	0.4535	0.3709
WHIRLPOOL CORP	3.6573	1.3554	0.3706
JOHNSON & JOHNSON	3.3391	1.2370	0.3705
UNION PACIFIC CORP	3.9976	1.4787	0.3699
EMCO LTD	1.2640	0.4675	0.3699
BRENCO INC	1.0719	0.3956	0.3691
QUAKER CHEMICAL CORP	1.8208	0.6697	0.3678
ENNIS INC	1.5412	0.5663	0.3675
PENN ENGINEERING & MFG CORP	2.1746	0.7990	0.3674
FLEETWOOD ENTERPRISES INC	1.2860	0.4721	0.3671
MEAD CORP	2.8226	1.0360	0.3670
C S X CORP	3.5279	1.2926	0.3664
STARRETT L S CO	2.2072	0.8069	0.3656
HASTINGS MANUFACTURING CO	1.4691	0.5357	0.3646
OMNICOM GROUP INC	2.3829	0.8688	0.3646
ALCOA INC	3.5798	1.3048	0.3645
MASCO CORP	1.4129	0.5149	0.3644
FLEXSTEEL INDUSTRIES INC	1.2336	0.4495	0.3644
SONOCO PRODUCTS CO	2.4434	0.8889	0.3638
FORD MOTOR CO DEL	5.6620	2.0592	0.3637
WORTHINGTON INDUSTRIES INC	1.3064	0.4740	0.3628
WILEY JOHN & SONS INC	1.9973	0.7246	0.3628
ASARCO INC	2.7587	0.9979	0.3617
MARTIN MARIETTA CORP NEW	3.8269	1.3837	0.3616
DEERE & CO	3.6271	1.3105	0.3613
LUBRIZOL CORP	3.0186	1.0894	0.3609
MAPCO INC	2.8163	1.0152	0.3605
MODINE MANUFACTURING CO	2.7263	0.9811	0.3598
NORTHROP GRUMMAN CORP	3.8260	1.3731	0.3589
SCHAWK INC	0.6577	0.2357	0.3584
ECOLAB INC	1.6463	0.5892	0.3579
A B M INDUSTRIES INC	1.5860	0.5656	0.3566
V F CORP	3.5081	1.2470	0.3555
HONEYWELL INC	5.0125	1.7800	0.3551
LEE ENTERPRISES INC	1.8075	0.6400	0.3541
WETTERAU INC	1.7930	0.6341	0.3537

(continued)

Company	EPS	DPS	Payout ratio
MANITOWOC CO INC	1.9133	0.6750	0.3528
HARLAND JOHN H CO	1.6208	0.5717	0.3527
CLARCOR INC	2.1400	0.7530	0.3519
SIEBERT FINANCIAL CORP	0.9191	0.3231	0.3515
HITACHI LIMITED	2.1526	0.7530	0.3498
MARATHON OIL CORP	3.5127	1.2278	0.3495
FEDERAL SCREW WKS	2.7365	0.9561	0.3494
AMPCO PITTSBURGH CORP	1.1924	0.4160	0.3489
MACMILLAN BLOEDEL LTD	1.3955	0.4860	0.3482
LANCASTER COLONY CORP	2.1375	0.7426	0.3474
AMERICAN BUSINESS PRODS INC GA	1.5313	0.5312	0.3469
UNITED INDUSTRIAL CORP	1.4455	0.5011	0.3467
ASHLAND INC NEW	3.7363	1.2946	0.3465
FEDERAL PAPER BOARD INC	2.7756	0.9592	0.3456
SUPERIOR UNIFORM GROUP INC	1.3300	0.4591	0.3452
RAVEN INDUSTRIES INC	1.3751	0.4742	0.3449
HALLIBURTON COMPANY	2.6857	0.9262	0.3449
PACCAR INC	5.3748	1.8527	0.3447
LIMITED BRANDS INC	1.5413	0.5312	0.3446
LONGS DRUG STORES INC	2.1048	0.7247	0.3443
ONEIDA LTD	1.6130	0.5549	0.3440
C B I INDUSTRIES INC	3.3988	1.1682	0.3437
WILLAMETTE INDUSTRIES INC	3.0783	1.0554	0.3428
BAKER HUGHES INC	1.5472	0.5289	0.3418
KENNAMETAL INC	2.4287	0.8282	0.3410
C V S CAREMARK CORP	2.5880	0.8823	0.3409
STANDEX INTERNATIONAL CORP	1.8985	0.6456	0.3401
ANHEUSER BUSCH COS INC	2.8284	0.9566	0.3382
KELLY SERVICES INC	2.0805	0.7023	0.3375
BLACK & DECKER CORP	1.9409	0.6545	0.3372
CORNING INC	3.8714	1.3049	0.3370
AMERON INTERNATIONAL CORP DEL	3.5293	1.1890	0.3369
BUTLER MANUFACTURING CO DE	2.6450	0.8885	0.3359
IKON OFFICE SOLUTIONS INC	1.9153	0.6428	0.3356
CARLYLE INDUSTRIES INC	1.2187	0.4089	0.3355
ELECTRONIC DATA SYS CORP NEW	1.6503	0.5528	0.3350
N C H CORP	2.8249	0.9461	0.3349
HILL ROM HOLDINGS INC	2.1951	0.7350	0.3348
CHAMPION INTERNATIONAL CORP	2.2051	0.7380	0.3347
DRESSER INDUSTRIES INC	2.5825	0.8618	0.3337
APPLERA CORP	1.3165	0.4392	0.3336
PITT DES MOINES INC	2.9909	0.9971	0.3334
MURPHY OIL CORP	2.8769	0.9568	0.3326
KELLWOOD COMPANY	2.0543	0.6820	0.3320

(continued)

Company	EPS	DPS	Payout ratio
REYNOLDS & REYNOLDS CO	1.8778	0.6233	0.3319
CONSTAR INTERNATIONAL INC NEW	1.8253	0.6057	0.3318
C T S CORP	1.4677	0.4870	0.3318
CAROLINA FREIGHT CORP	1.2747	0.4220	0.3311
PULSE ELECTRONICS CORP	1.2926	0.4270	0.3304
ALICO INC	1.2578	0.4151	0.3300
CLIFFS NATURAL RESOURCES INC	4.4192	1.4543	0.3291
RESEARCH INC	0.9123	0.2995	0.3282
ROHM & HAAS CO	4.1020	1.3450	0.3279
MCCORMICK & CO INC	1.9843	0.6500	0.3276
GLATFELTER P H CO	2.6174	0.8553	0.3268
UNOCAL CORP	3.3820	1.1040	0.3264
TEXAS INSTRUMENTS INC	2.6226	0.8530	0.3252
GREY GLOBAL GROUP INC	7.7818	2.5256	0.3246
Y R C WORLDWIDE INC	2.0988	0.6803	0.3241
PERKINELMER INC	1.2435	0.4029	0.3240
JOHNSON CONTROLS INC	3.4451	1.1154	0.3238
LEGGETT & PLATT INC	1.6663	0.5387	0.3233
INTERPUBLIC GROUP COS INC	2.6687	0.8618	0.3229
SCHLUMBERGER LTD	2.9066	0.9365	0.3222
HARRIS CORP	2.5954	0.8335	0.3211
BRIDGFORD FOODS CORP	0.6023	0.1932	0.3208
ALBERTSONS INC	1.9968	0.6394	0.3202
AMETEK INC NEW	1.8592	0.5944	0.3197
CRANE CO	2.9520	0.9433	0.3195
UNIVAR CORP	1.5540	0.4960	0.3192
AMERICAN MAIZE PRODS CO	1.5556	0.4961	0.3189
SHERWIN WILLIAMS CO	3.1211	0.9937	0.3184
CALMAT CO	2.6650	0.8483	0.3183
LOCKHEED MARTIN CORP	4.6655	1.4770	0.3166
CORE INDUSTRIES INC	1.3761	0.4347	0.3159
KNIGHT RIDDER INC	2.8792	0.9081	0.3154
CON WAY INC	2.3073	0.7259	0.3146
KOLLMORGEN CORP	0.9335	0.2931	0.3140
DEAN FOODS CO	2.3313	0.7314	0.3137
MATTEL INC	1.1671	0.3661	0.3137
AMERICAN GREETINGS CORP	1.6226	0.5089	0.3136
RUBBERMAID INC	1.7181	0.5367	0.3124
O SULLIVAN CORP	1.3439	0.4192	0.3119
CORDANT TECHNOLOGIES INC	2.7068	0.8409	0.3107
PARKER HANNIFIN CORP	2.9837	0.9268	0.3106
RAYTHEON CO	3.9104	1.2146	0.3106
PREMIER INDUSTRIAL CORP	1.6544	0.5130	0.3101
VALSPAR CORP	1.4568	0.4507	0.3094

(continued)

Company	EPS	DPS	Payout ratio
STANDARD MOTOR PRODUCTS INC	1.2109	0.3741	0.3089
GREAT NORTHERN NEKOOSA CORP	4.6923	1.4476	0.3085
REGAL BELOIT CORP	1.6186	0.4987	0.3081
PALL CORP	1.4568	0.4470	0.3068
SMITH A O CORP	2.3277	0.7135	0.3065
METHODE ELECTRONICS INC	0.6966	0.2131	0.3060
DOLE FOOD INC NEW	1.6069	0.4915	0.3058
CARLISLE COMPANIES	2.8640	0.8713	0.3042
PULASKI FURNITURE CORP	1.6240	0.4919	0.3029
BALL CORP	2.5113	0.7600	0.3026
LEARONAL INC	1.4447	0.4353	0.3013
LAWSON PRODUCTS INC	1.4305	0.4300	0.3006
STANDARD PRODUCTS CO	2.7506	0.8263	0.3004
GRACO INC	1.9968	0.5972	0.2990
BOEING CO	3.6459	1.0902	0.2990
RALSTON PURINA CO	2.6094	0.7800	0.2989
BALDOR ELECTRIC CO	1.3851	0.4116	0.2972
VILLAGE SUPER MARKET INC	2.1813	0.6478	0.2970
RYDER SYSTEMS INC	2.3212	0.6892	0.2969
HUNT CORP	1.1890	0.3529	0.2968
APPLIED INDUSTRIAL TECHS INC	2.1677	0.6409	0.2957
H N I CORP	1.7347	0.5127	0.2956
KAMAN CORP	1.7440	0.5146	0.2950
OHIO ART CO	1.0906	0.3203	0.2937
MAGNA INTERNATIONAL INC	3.0680	0.9010	0.2937
TELEFLEX INC	2.1546	0.6306	0.2927
DOMTAR INC	1.7595	0.5144	0.2923
CHURCH & DWIGHT INC	1.9126	0.5555	0.2905
CABOT CORP	2.8130	0.8157	0.2900
QUANEX CORP	2.5367	0.7353	0.2899
WEST PHARMACEUTICAL SERVICES INC	1.5165	0.4382	0.2890
TRINITY INDUSTRIES INC	1.8938	0.5465	0.2886
GIANT FOOD INC	2.5452	0.7341	0.2884
WALGREEN CO	2.1827	0.6288	0.2881
ALLIANCE ONE INTERNATIONAL INC	2.7500	0.7850	0.2855
HARCOURT GENERAL INC	1.9674	0.5608	0.2850
AIR PRODUCTS & CHEMICALS INC	3.1108	0.8866	0.2850
CARTER WALLACE INC	1.5206	0.4321	0.2842
HAVERTY FURNITURE COS INC	1.3445	0.3821	0.2842
GOODHEART WILLCOX INC	2.4671	0.7000	0.2837
TEKTRONIX INC	2.0470	0.5795	0.2831
RITE AID CORP	1.6887	0.4780	0.2831
FLOWSERVE CORP	2.3354	0.6606	0.2829
NACCO INDUSTRIES INC	3.4328	0.9701	0.2826

(continued)

Company	EPS	DPS	Payout ratio
WOODWARD INC	12.0189	3.3700	0.2804
FLUOR CORP NEW	2.1162	0.5919	0.2797
NEW YORK TIMES CO	2.1436	0.5994	0.2796
GRAINGER W W INC	3.2007	0.8926	0.2789
NORDSON CORP	2.1639	0.6004	0.2774
OWENS & MINOR INC NEW	1.3859	0.3818	0.2755
BANDAG INC	3.4447	0.9488	0.2754
BECTON DICKINSON & CO	2.7480	0.7552	0.2748
KIMBALL INTERNATIONAL INC	2.1268	0.5843	0.2747
MEDIA GENERAL INC	2.3121	0.6339	0.2742
IMPERIAL OIL LTD	2.7651	0.7553	0.2731
AUTOMATIC DATA PROCESSING INC	1.9911	0.5410	0.2717
DOVER CORP	2.9127	0.7904	0.2714
MET PRO CORP	0.8377	0.2272	0.2712
STEPAN CO	2.4189	0.6549	0.2707
MACDERMID INC	1.6949	0.4577	0.2701
ILLINOIS TOOL WORKS INC	2.8676	0.7737	0.2698
BOB EVANS FARMS INC	1.4688	0.3957	0.2694
SEAGRAM LTD	3.4437	0.9260	0.2689
WALLACE COMPUTER SERVICES INC	2.1295	0.5664	0.2660
MOLSON COORS BREWING CO	2.1654	0.5747	0.2654
HORMEL FOODS CORP	2.5758	0.6833	0.2653
TARGET CORP	3.1654	0.8367	0.2643
TRANS LUX CORP	0.5265	0.1389	0.2638
MEREDITH CORP	2.7666	0.7298	0.2638
APOGEE ENTERPRISES INC	0.7708	0.2016	0.2615
CASCADE CORP	2.6393	0.6896	0.2613
DELTA AIR LINES INC	2.6118	0.6803	0.2605
CANADIAN PACIFIC RAILWAY LTD	2.8130	0.7300	0.2595
SEAWAY FOOD TOWN INC	1.7928	0.4615	0.2574
LA Z BOY INC	2.3243	0.5957	0.2563
JOY GLOBAL INC	2.5851	0.6594	0.2551
BANTA CORP	2.0394	0.5201	0.2550
COHU INC	0.8212	0.2086	0.2540
RUSSELL CORP	1.4947	0.3779	0.2528
COURIER CORP	1.7486	0.4417	0.2526
TIDEWATER INC	2.5485	0.6428	0.2522
TEXAS INDUSTRIES INC	2.1431	0.5397	0.2518
TECK RESOURCES LTD	1.1536	0.2898	0.2512
BAIRNCO CORP	1.3779	0.3453	0.2506
JORGENSEN EARLE M CO DE NEW	3.4458	0.8578	0.2489
WEYCO GROUP INC	3.2289	0.8034	0.2488
COOPER TIRE & RUBBER CO	1.8396	0.4577	0.2488
T D K CORP	2.2116	0.5501	0.2487

(continued)

Company	EPS	DPS	Payout ratio
SONY CORP	1.0057	0.2501	0.2487
WASHINGTON POST CO	14.3671	3.5656	0.2482
BECKMAN COULTER INC	1.6713	0.4116	0.2463
CUBIC CORP	1.6085	0.3925	0.2440
BARD C R INC	2.0320	0.4951	0.2437
FULLER H B CO	1.8451	0.4494	0.2435
WENDYS INTERNATIONAL INC	1.0976	0.2668	0.2431
MITSUI & CO LTD	6.8629	1.6680	0.2430
GREIF INC	2.6358	0.6405	0.2430
SUPERIOR INDUSTRIES INTL INC	1.3535	0.3277	0.2421
AVNET INC	2.2024	0.5324	0.2417
FAMILY DOLLAR STORES INC	1.3566	0.3265	0.2407
COMMERCIAL INTERTECH CORP	2.5239	0.6064	0.2402
WACOAL HOLDINGS CORP	1.5532	0.3724	0.2397
FRANKLIN ELECTRIC INC	2.1174	0.5037	0.2379
SIFCO INDUSTRIES INC	1.0563	0.2509	0.2375
SCHULMAN A INC	1.8767	0.4433	0.2362
ROBBINS & MYERS INC	2.1030	0.4961	0.2359
MILLIPORE CORP	1.3730	0.3228	0.2351
FRISCHS RESTAURANTS INC	1.3584	0.3190	0.2348
MCDONNELL DOUGLAS CORP	5.1794	1.2126	0.2341
AMERICAN STORES CO NEW	2.8476	0.6655	0.2337
NEWMARKET CORP	3.8581	0.8886	0.2303
EDO CORP	1.0527	0.2422	0.2301
TYCO INTERNATIONAL LTD SWTZLND	1.9844	0.4560	0.2298
PEP BOYS MANNY MOE & JACK	1.8804	0.4318	0.2296
CANON INC	2.2080	0.5069	0.2296
MAKITA CORP	1.4286	0.3250	0.2275
WAUSAU PAPER CORP	1.5340	0.3482	0.2270
FARMER BROTHERS CO	5.2408	1.1865	0.2264
SYSCO CORP	1.7452	0.3946	0.2261
SERVICE CORP INTERNATIONAL	1.4043	0.3155	0.2247
DONALDSON INC	2.0112	0.4515	0.2245
BOWNE & CO INC	1.1956	0.2674	0.2236
MOTOROLA SOLUTIONS INC	2.8209	0.6304	0.2235
MINE SAFETY APPLIANCES CO	3.9163	0.8739	0.2231
T J X COMPANIES INC NEW	2.0726	0.4603	0.2221
RUBY TUESDAY INC	1.5098	0.3349	0.2219
DOUGLAS & LOMASON CO	1.8438	0.4080	0.2213
KYOCERA CORP	3.0711	0.6650	0.2165
MERCANTILE STORES INC	4.8994	1.0570	0.2157
MYERS INDUSTRIES INC	1.0405	0.2229	0.2143
KANSAS CITY SOUTHERN	3.3614	0.7169	0.2133
HUFFY CORP	1.9209	0.4070	0.2119

(continued)

Company	EPS	DPS	Payout ratio
NOLAND COMPANY	1.9859	0.4116	0.2072
ROWE COS	1.2036	0.2489	0.2068
MILLER HERMAN INC	1.6546	0.3390	0.2049
ACETO CORP	1.2322	0.2517	0.2042
WOLVERINE WORLD WIDE INC	1.2032	0.2443	0.2031
TOOTSIE ROLL INDS INC	1.5992	0.3233	0.2022
TORO COMPANY	2.4127	0.4877	0.2021
C A E INC	0.7205	0.1455	0.2019
PLACER DOME INC	1.0794	0.2159	0.2000
NEXEN INC	1.6830	0.3360	0.1997
TRANZONIC COMPANIES	1.2284	0.2429	0.1977
AGILYSYS INC	0.7226	0.1422	0.1969
NORDSTROM INC	2.0356	0.3993	0.1961
MOSINEE PAPER CORP	1.9169	0.3745	0.1954
PANASONIC CORP	2.7438	0.5327	0.1941
MEDTRONIC INC	2.9024	0.5516	0.1900
NOVO NORDISK A S	2.9530	0.5584	0.1891
DANIEL INDUSTRIES INC	1.0190	0.1925	0.1889
COMMERCIAL METALS CO	2.3413	0.4418	0.1887
GAP INC	1.6536	0.3119	0.1886
RYLAND GROUP INC	2.1292	0.3997	0.1877
P V H CORP	1.5833	0.2953	0.1865
M T S SYSTEMS CORP	1.6641	0.3063	0.1841
SEQUA CORP	2.7823	0.5117	0.1839
ALLEN ORGAN CO	2.3957	0.4385	0.1831
HESS CORP	3.6960	0.6748	0.1826
KEITHLEY INSTRUMENTS INC	0.8238	0.1493	0.1812
G & K SERVICES INC	1.1816	0.2138	0.1809
WAL MART STORES INC	1.9279	0.3458	0.1794
VALMONT INDUSTRIES INC	2.2791	0.4063	0.1783
DOLLAR GENERAL CORP NEW	1.1139	0.1960	0.1760
SIGMA ALDRICH CORP	2.4511	0.4278	0.1746
LOWES COMPANIES INC	1.6152	0.2816	0.1744
WARWICK VALLEY TELEPHONE CO	3.7775	0.6405	0.1696
GREAT LAKES CHEM CORP	2.3097	0.3890	0.1684
ARCHER DANIELS MIDLAND CO	1.9606	0.3260	0.1663
IPSCO INC	2.7394	0.4339	0.1584
APACHE CORP	2.3063	0.3594	0.1558
HELMERICH & PAYNE INC	2.0855	0.3249	0.1558
DISNEY WALT CO	2.6260	0.4073	0.1551
PIONEER CORP JAPAN	1.1861	0.1703	0.1435
HEWLETT PACKARD CO	2.6385	0.3753	0.1422
NOBLE ENERGY INC	1.6858	0.2297	0.1363
HONDA MOTOR LTD	2.9264	0.3884	0.1327

(continued)

Company	EPS	DPS	Payout ratio
CRACKER BARREL OLD COUNTRY STORE	1.4434	0.1746	0.1210
V S E CORP	1.6139	0.1906	0.1181
STANDARD COMMERCIAL CORP	2.9253	0.3358	0.1148
SHENANDOAH TELECOM COMPANY	3.4106	0.3850	0.1129
FUJIFILM HOLDINGS CORP	2.0113	0.2241	0.1114
SEA CONTAINERS LTD	4.6123	0.4984	0.1081
TYSON FOODS INC	1.3272	0.1416	0.1067
VIRCO MFG CORP	0.7891	0.0740	0.0938
CENTEX CORP	2.1800	0.2016	0.0925
BRINKS CO	2.1083	0.1934	0.0917
DILLARDS INC	2.6262	0.2214	0.0843
HEICO CORP NEW	1.0943	0.0919	0.0839
CIRCUIT CITY STORES INC	1.2505	0.1037	0.0829
INTERNATIONAL SPEEDWAY CORP	2.0574	0.1649	0.0801
UNIFIRST CORP	1.8493	0.1405	0.0760
COMINCO LTD	1.9819	0.1492	0.0753
SOUTHWEST AIRLINES CO	1.5074	0.0945	0.0627
PRECISION CASTPARTS CORP	2.8621	0.1393	0.0487
C T COMMUNICATIONS INC	13.1824	0.2511	0.0190
JOURNAL COMMUNICATIONS INC	3.0321	0.0424	0.0140
DART GROUP CORP	-0.0200	0.1303	-6.5141
Mean	2.4290	0.9159	0.3636
Variance	1.8977	0.3768	0.0985
Standard deviation	1.3776	0.6139	0.3139
Kurtosis	20.6265	4.7306	380.6515
Skewness	3.3799	1.7729	-17.2978

References

- Almeida, H., M. Campello, and A.F. Galvao Jr. 2010. Measurement errors in investment equations. *Review of Financial Studies* 23: 3279–3328.
- Ang, J.S. 1975. Dividend policy: Informational content or partial adjustment? *Review of Economics and Statistics* 57: 65–70.
- Chen, H.Y., M.C. Gupta, A.C. Lee, and C.F. Lee. 2013. Sustainable growth rate, optimal growth rate, and optimal payout ratio: A joint optimization approach. *Journal of Banking & Finance* 37: 1205–1222.
- Darby, M.R. 1972. The allocation of transitory income among consumers' assets. *American Economic Review* (September): 928–941.
- . 1974. The permanent income theory of consumption—A restatement. *Quarterly Journal of Economic* (May): 228–250.
- Dichev, I.D., and V.W. Tang. 2009. Earnings volatility and earnings predictability. *Journal of Accounting and Economics* 47: 160–181.
- Eisner, R. 1967. A permanent income theory of investment. *American Economic Review* 57: 363–390.

- . 1978. *Factors in business investment*. General Series No. 102. Washington, DC: National Bureau of Economic Research.
- Fama, E.F., and H. Blasiak. 1968. Dividend policy: An empirical analysis. *Journal of American Statistical Association* 63: 1132–1161.
- Frankel, R., and L. Litov. 2009. Earnings persistence. *Journal of Accounting and Economics* 47: 182–190.
- Friedman, M. 1957. *A theory of the consumption function*. Princeton, NJ: Princeton University Press.
- Garrett, I., and R. Priestley. 2000. Dividend behavior and dividend signaling. *The Journal of Financial and Quantitative Analysis* 35 (June): 173–189.
- Greene, W.H., 1993. *Econometric Analysis*. Macmillan Publishing Company, New York.
- Ham, C.G., Z.R. Kaplan, and M.T. Leary. 2020. Do dividends convey information about future earnings? *Journal of Financial Economics* 136 (2): 547–570.
- Higgins, R. 1977. How much growth can a firm afford? *Financial Management* 6: 7–16.
- Higgins, R.C. 1981. Sustainable growth under inflation. *Financial Management* 10: 36–40.
- . 2008. *Analysis for financial management*. 9th ed. New York: McGraw-Hill, Inc.
- Kmenta, J. 1986. *Elements of econometrics*. 2nd ed. New York: Macmillan.
- Kumar, P., and B.S. Lee. 2001. Discrete dividend policy with permanent earnings. *Financial Management*: 55–76.
- Lambrecht, B.M., and S.C. Myers. 2012. A Lintner Model of Payout and Managerial Rents. *Journal of Finance* 67 (October): 1761–1810.
- Latane, H.A., and C.P. Jones. 1979. Standardized unexpected earnings-1971–77. *Journal of Finance* 34: 717–724.
- Lee, C.F., and W.J. Primeaux. 1991. Current- versus permanent- dividend payments behavioral model: Methods and applications. *Advances in Quantitative Analysis of Finance and Accounting* 1 (Part A): 109–130.
- Lee, C.F., M. Djarraya, and C. Wu. 1987. A further empirical investigation of the dividend adjustment process. *Journal of Econometrics*: 267–285.
- Lee, C.F., M.C. Gupta, H.Y. Chen, and A.C. Lee. 2011. Optimal payout ratio under uncertainty and the flexibility hypothesis: Theory and empirical evidence. *Journal of Corporate Finance* 17: 483–501.
- Lee, C.F., H.Y. Chen, and A.C. Lee. 2015. Alternative errors-in-variables models and their applications in finance research. *The Quarterly Review of Economics and Finance* 58: 213–227. working paper.
- Lee, C.F., H.Y. Chen, and J.C. Lee. 2019. *Financial econometrics, mathematics and statistics*. Springer Academic Publishers.
- Lintner, J. 1956. Distribution of income of corporations among dividends, retained earnings, and taxes. *American Economic Review* (May): 97–113.
- Marsh, T.A., and R.C. Merton. 1987. Dividend behavior for the aggregate stock market. *The Journal of Business* 60 (January): 1–40.
- Miller, M.H., and F. Modigliani. 1961. Dividend policy, growth and valuation of shares. *Journal of Business* 34 (October): 411–433.
- . 1966. Some estimates of the cost of capital to the electric utility industry. *American Economic Review* 56: 334–391.
- Miller, M.H., and M.S. Scholes. 1982. Dividends and taxes: Some empirical evidence. *Journal of Political Economy* 90: 1118–1141.
- Modigliani, F., and M.H. Miller. 1958. The cost of capital, corporation finance and the theory of investment. *American Economic Review* 4B: 261–297.
- . 1963. Corporate income tax and the cost of capital: A correction. *American Economic Review* 53: 433–443.
- Pan, M.S. 2001. Aggregate dividend behavior and permanent earnings hypothesis. *Financial Review* 36 (1): 23–38.
- Wang, N. 2003. Caballero meets Bewley: The permanent-income hypothesis in general equilibrium. *American Economic Review* 93: 927–936.
- Weston, F.J., E. Brigham, and S. Besley. 2004. *Essentials of managerial finance*. Cengage South-Western.



Michael Dothan

Abstract

Assessing financing effects, such as tax benefits of debt, is an important part of financial analysis and valuation. This chapter describes the frameworks and valuation formulas of Modigliani-Miller, Miles-Ezzell, modified Miles-Ezzell, Modigliani-Miller with growth, Fernandez, and other related setups. In addition, the chapter examines recent dynamic models that value interest tax shields in the context of flexible financing strategies. The chapter starts with two summary tables that list notation, definitions, and formulas for some discrete-time valuation models. Because of the diversity of views on the appropriate discount rate for interest tax shields, I present a short derivation of the Miles-Ezzell formula. The last part of the chapter reviews recent dynamic valuation models of interest tax shields, including the application of barrier options to analyzing the effects of flexible financing strategies, and presents a summary that compares the valuation outcomes of the various models.

Keywords

Barrier options · Default · Financing strategies · Leverage · Tax shields · Valuation

M. Dothan (✉)

Atkinson Graduate School of Management, Willamette University, Salem, OR, USA

e-mail: mdothan@willamette.edu

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,

https://doi.org/10.1007/978-3-030-91231-4_71

1641

71.1 Introduction

Enterprise value of a leveraged firm is the sum of equity value and debt value. In standard financial models, enterprise value can be computed either by discounting unleveraged free cash flows at the weighted average cost of capital (WACC), or by adding the effects of financing strategies, such as the value of interest tax shields, to the unleveraged firm value that represents operating decisions. Interest tax shields are an important part of enterprise value, Graham (2000) estimates the tax benefits of debt realized by a typical firm at 9.7% of enterprise value, and Kemsley and Nissim (2002) estimate that the average value of interest tax shields is 10% of enterprise value. For high-debt firms, acquired by private equity firms in a leveraged buyout, Kaplan and Strömberg (2009) estimate the value of interest tax shields at 10–20% of enterprise value.

The process of separating operating and financing effects, proposed by Modigliani and Miller (1963) and Myers (1974), is known as the adjusted present value (APV) method of valuation. Luehrman (1997) promotes the view that APV is the better alternative. Whereas WACC requires only one discounting operation rather than two or more, it is not as good as APV at taking care of the effects of financing strategies. In addition, APV is better at unbundling and making visible the components of value, and can be customized according to specific circumstances and managers' needs. The popular corporate finance textbook Brealey et al. (2011) echoes those sentiments, saying that the APV method "gives the financial manager an explicit view of the factors that are adding or subtracting value." Koller et al. (2010), through several editions of their advanced book on valuation, are also strong proponents of separating the effects of operating and financing strategies.

Booth (2002) contends that the APV method is not always consistent with the WACC method, and demurs at the assumption of constant market leverage that characterizes the so-called Miles-Ezzell formula, presented here in Sect. 71.3.3. I will argue in Sect. 71.5 that recently developed valuation models with dynamic debt policies allow us to attenuate those concerns.

The recent models exploit a link between barrier options and the value of interest tax shields. Barrier options come to life, expire, and/or make a cash payoff contingent on a barrier event, such as EBIT hitting a specified barrier for the first time. The link between barrier options and interest tax shields arises in explicit modeling of default and refinancing, with the associated loss or increase of interest tax shields.

I start this chapter with two summary tables that list notation, definitions, and formulas for some discrete-time valuation models. Next, I describe two influential models of interest tax shields: Modigliani-Miller and Miles-Ezzell. Because of the diversity of views on the appropriate discount rate for interest tax shields, I present a short derivation of the Miles-Ezzell formula. Subsequently, I consider several alternative models with fixed debt or leverage ratio. The last part of this chapter reviews recent dynamic valuation models of interest tax shields, including the application of barrier options to analyzing the effects of flexible financing strategies, and presents a summary that compares the valuation outcomes of the various models.

Table 1 Notation, definitions, and formulas for discrete-time valuation models

1	Market equity value	S_t
2	Market debt value	D_t
3	Market leveraged enterprise value	$W_t = S_t + D_t$
4	Market unleveraged firm value	V_t
5	Coupon or interest expense	C_{t+1}
6	Corporate tax rate	T_C
7	EBIT	X_{t+1}
8	Invested capital	U_t
9	Return on invested capital (ROIC)	$R_{t+1} = \frac{(1-T_C)X_{t+1}}{U_t}$
10	Unleveraged free cash flow	$FCF_{t+1} = (1 - T_C)X_{t+1} - (U_{t+1} - U_t)$
11	Leveraged free cash flow	$FCF_{t+1}^* = FCF_{t+1} + T_C C_{t+1}$
12	Leveraged equity cash flow	$ECF_{t+1}^* = FCF_{t+1} - (1 - T_C)C_{t+1} + D_{t+1} - D_t$
13	Risk-free or cash rate	r_F
14	Conditioning information	$\mathcal{F}_t$
15	Required equity return or cost of equity	$r_{S,t} = \frac{E(FCF_{t+1}^* + S_{t+1} \mathcal{F}_t) - S_t}{S_t}$
16	Cost of debt	$r_{D,t} = \frac{E(C_{t+1} \mathcal{F}_t)}{D_t}$
17	Market leverage ratio	$L_t = \frac{D_t}{S_t}$
18	Weighted average cost of capital (WACC)	$r_{W,t} = \frac{E(FCF_{t+1} + W_{t+1} \mathcal{F}_t) - W_t}{W_t}$
19	Formula for weighted average cost of capital (WACC)	$r_{W,t} = \frac{r_{S,t} + (1 - T_C)L_t r_{D,t}}{1 + L_t}$
20	Required asset return or cost of unleveraged equity	$r_{A,t} = \frac{E(FCF_{t+1} + V_{t+1} \mathcal{F}_t) - V_t}{V_t}$
21	Pricing kernel	z_{t+1}
22	Risk-free discount factor	$\frac{1}{1+r_F} = E(z_{t+1} \mathcal{F}_t)$
23	Market unleveraged firm value using pricing kernel	$V_t = E[(FCF_{t+1} + V_{t+1})z_{t+1} \mathcal{F}_t]$
24	Market leveraged enterprise value using pricing kernel	$W_t = E[(FCF_{t+1}^* + W_{t+1})z_{t+1} \mathcal{F}_t]$

71.2 Notation, Definitions, and Formulas

The items in Tables 1 and 2 refer to a perpetual model. Time index t is an integer $t \geq 0$. Some table entries are indexed by $t + 1$ rather than t for convenience, because that is how they appear in the formulas. Regardless of the notation, all such entries are defined for every discrete-time starting at 0.

It is well-known that the weighted average cost of capital in line 19 of Table 1 is the suitable discount rate for the unleveraged free cash flow FCF_{t+1} to obtain leveraged enterprise value at time t , as implied by line 18 of Table 1. The calculation

Table 2 Notation, definitions, and formulas for discrete-time stationary valuation models

1	Fixed mean growth	g
2	Fixed market leverage ratio	$L_t = \frac{D_t}{S_t}$
3	Market unleveraged firm value	$V_t = \frac{FCF_t(1+g)}{r_A - g}$
4	Market leveraged enterprise value	$W_t = \frac{FCF_t(1+g)}{r_W - g}$
5	Adjustment factor	$\lambda = \frac{T_C r_D}{1+r_D}$
6	Formula for WACC in terms of required asset return	$1 + r_W = \left(1 - \frac{\lambda L}{1+L}\right)(1 + r_A)$
7	Formula for required asset return	$r_A = \frac{r_S + (1-\lambda)Lr_D}{1 + (1-\lambda)L}$
8	Valuation of interest tax shields	$VTS_t = \frac{T_C r_D D_t}{r_A - g} \frac{1+r_A}{1+r_D}$

is straightforward, multiply line 15 by $\frac{1}{1+L_t}$, multiply line 16 by $\frac{(1-T_C)L_t}{1+L_t}$, and add, then use lines 12 and 18 to get line 19.

The technical meaning of conditioning information $\mathcal{F}_t$ in line 14 is the σ -field of events that are decidable at time t . Coupon payment in line 5 is random, and reflects the possibility of default. The pricing kernel in line 21 of Table 1 adjusts future cash flows and values for risk-neutral valuation. In particular, the conditional expectation of the pricing kernel is the risk-free discount factor, as shown in line 22.

The stationary or steady-state model in Table 2 is characterized by a constant mean growth rate g and fixed market leverage ratio L . Fixed market leverage ratio means that, as enterprise value evolves over time, the amount of debt is rebalanced or adjusted at each discrete-time point to maintain a fixed ratio of market debt value to market equity value, or, equivalently, of market debt value to market enterprise value.

The formulas in lines 3 and 4 of Table 2 represent standard valuation of perpetual growth annuities. The significance of the adjustment factor in line 5 is due to its part in the relationship between WACC and required asset return in line 6, which leads to the valuation formula of interest tax shields in line 8, as shown in Sect. 71.3.3. Lastly, it is important to recognize that notation in papers cited in this chapter varies, and, in particular, some authors define the market leverage ratio as $\frac{D_t}{W_t}$, which corresponds to $\frac{L_t}{1+L_t}$ in our notation.

71.3 Two Influential Models

Two influential works that analyze the value of interest tax shields are the combination of Modigliani and Miller (1958, 1963), and the combination of Miles and Ezzell (1980, 1985).

71.3.1 Modigliani-Miller Formula

The Modigliani and Miller (1958, 1963) model is set in discrete time, debt is perpetual, the value of debt is fixed, and the cost of debt is constant. This framework

corresponds to the unleveraged free cash flow fluctuating around a constant level $FCF_{t+1} = FCF_t(1 + \varepsilon_{t+1})$, with $E(\varepsilon_{t+1} | \mathcal{F}_t) = 0$, and $D_t = D_0$. It is important to recognize that the model assumes that interest expense C_t is non-random, so that $r_{D,t} = \frac{C_{t+1}}{D_t} = \frac{C_{t+1}}{D_0}$ is a constant $r_{D,t} = r_D$. The model does not deduct the value of deadweight default cost. Starting with the recursive equation for a perpetual annuity

$$VTS_t = \frac{T_C r_D D_0}{1 + r_D} + \frac{VTS_t}{1 + r_D} \quad (1)$$

and solving for VTS_t , we get

$$VTS_t = \frac{T_C r_D D_0}{r_D} = T_C D_0 \quad (2)$$

Notably, Eq. (2) is valid only if $r_D > 0$. Otherwise, if $r_D = 0$, Eq. (1) is a tautology, and has nothing to say about the value of interest tax shields. We can, however, use general financial considerations to conclude that an interest-free loan generates zero interest tax shields. Therefore, a more precise form of Eq. (2) is

$$VTS_t = \begin{cases} T_C D_0 & \text{if } r_D > 0 \\ 0 & \text{if } r_D = 0 \end{cases} \quad (3)$$

I return to this issue in Sect. 71.4.

71.3.2 Example

The firm's fixed debt is $D_t = D_0 = 250$, constant cost of debt is $r_D = 8\%$, and the corporate tax rate is $T_C = 25\%$. Using Eq. (3) we get $VTS_t = 62.50$.

71.3.3 Miles-Ezzell Formula

The model of Miles and Ezzell (1980, 1985) is set in discrete time, debt is perpetual, and market leverage ratio is fixed. This framework corresponds to the stationary model of unleveraged free cash flow fluctuating around a growing level $FCF_{t+1} = FCF_t(1 + g + \varepsilon_{t+1})$, with noise ε_t drawn from a fixed distribution, $E(\varepsilon_{t+1} | \mathcal{F}_t) = 0$, and constant market leverage $L_t = L_0$, implying that $r_{S,t} = r_S$, $r_{D,t} = r_D$, and $r_{A,t} = r_A$. The model does not deduct the value of deadweight default cost. The Miles-Ezzell formula is

$$VTS_t = \frac{T_C r_D D_t}{r_A - g} \frac{1 + r_A}{1 + r_D} \quad (4)$$

Equation (4) follows from the relationship between weighted average cost of capital and required asset return in line 6 of Table 2

$$1 + r_W = \left(1 - \frac{\lambda L}{1 + L}\right)(1 + r_A) \quad (5)$$

I present a short verification of the Miles-Ezzell formula from Eq. (5) and, in Sect. 71.3.5, offer a short derivation of Eq. (5) based on Arzac and Glosten [2005]. Combining Eq. (5) with the definition of the constant market leverage ratio in line 2 of Table 2, and multiplying by W_t implies

$$(r_W - r_A)W_t = -\lambda D_t(1 + r_A) \quad (6)$$

From line 0 of Table 2 we have

$$(r_W - g)W_t = FCF_t(1 + g) \quad (7)$$

Subtracting Eq. (6) from Eq. (7), dividing by $r_A - g$, and using the formula for the adjustment factor in line 5 of Table 2, yields

$$W_t = \frac{FCF_t(1 + g)}{r_A - g} + \frac{TCr_D D_t}{r_A - g} \frac{1 + r_A}{1 + r_D} \quad (8)$$

The first term on the right side of Eq. (8) is the value of the unleveraged firm, and the second term is the value of interest tax shields, as presented in Eq. (4). It is important to note that even with expected zero growth, the Modigliani-Miller and Miles-Ezzell valuation formulas produce different results. Indeed, substituting $g = 0$ in Eq. (4), we get

$$VTS_t = \frac{TCr_D D_t}{r_A} \frac{1 + r_A}{1 + r_D} \quad (9)$$

Because $r_D < r_A$, we get a value that is less than the Modigliani-Miller value of TCr_D . With the same assumption about zero mean growth, this difference is due to different assumptions about the cost of debt.

An important factor in the application of the Miles-Ezzell formula is the availability of required asset return r_A . To the extent that we accept the stationary model as a rough approximation to the reality of at least some firms, we can employ the formula in line 0 of Table 2 to estimate r_A , using the same inputs as in the calculation of the weighted average cost of capital (WACC) r_W . I will return to the realism of the stationary model and a proposed relaxation of some of its assumptions in Sect. 71.5.

An important feature of the Miles-Ezzell formula in Eq. (4) is that it satisfies the recursive valuation equation

$$VTS_t = \frac{TCr_D D_t}{1 + r_D} + \frac{1 + g}{1 + r_A} VTS_t \quad (10)$$

This validates the notion that next year's tax savings should be discounted at the cost of debt, whereas all subsequent tax savings should be discounted at the required asset return. The Miles-Ezzell formula is favored by Lewellen and Emery (1986), Taggart (1991), Arzac and Glosten (2005), and Brealey et al. (2011). Interestingly, Harris and Pringle (1985), Ruback (1995, 2002), Cooper and Nyborg (2006), and Farber et al. (2006) present instead a modified Miles-Ezzell formula

$$VTS_t = \frac{T_C r_D D_t}{r_A - g} \quad (11)$$

A possible reconciliation is offered by Taggart (1991), Arzac and Glosten (2005), and Brealey et al. (2011), who observe that the Miles-Ezzell formula is valid for discrete rebalancing of debt to maintain a fixed market leverage ratio, whereas the modified Miles-Ezzell formula is valid for continuous rebalancing of debt, with r_D , r_A , and g representing continuous-time cost of debt, required asset return, and mean growth, respectively.

71.3.4 Example

Debt value is $D_t = 250$, cost of debt is $r_D = 8\%$, mean growth rate is $g = 7\%$, required asset return is $r_A = 12\%$, and corporate tax rate is $T_C = 25\%$. Applying Eq. (4) we get $VTS_t = 103.70$. For comparison, using Eq. (11) to compute the modified Miles-Ezzell value we get $VTS_t = 100.00$. The value of interest tax shields is larger here than in the Modigliani-Miller Example because of different assumptions about the cost of debt and the presence of positive mean growth.

71.3.5 Weighted Average Cost of Capital and Required Asset Return

The preceding validation of the Miles-Ezzell formula rests on the relationship in Eq. (5). To derive Eq. (5), notice that from line 3 of Table 2 the unleveraged firm value V_{t+1} is proportional to the free cash flow FCF_{t+1} , with the proportionality constant $\frac{1+g}{r_A-g}$. Therefore

$$V_t = E \left[(FCF_{t+1} + V_{t+1}) z_{t+1} | \mathcal{F}_t \right] = \frac{1+r_A}{r_A-g} E(FCF_{t+1} z_{t+1} | \mathcal{F}_t) \quad (12)$$

Similarly, from line 4 of Table 2, the leveraged enterprise value W_{t+1} is proportional to unleveraged free cash flow FCF_{t+1} , with the proportionality constant $\frac{1+g}{r_W-g}$. Therefore, starting with leveraged free cash flow and using line 11 of Table 1

$$\begin{aligned}
W_t &= E[(FCF_{t+1}^* + W_{t+1})z_{t+1} | \mathcal{F}_t] \\
&= E[(FCF_{t+1} + T_C C_{t+1} + W_{t+1})z_{t+1} | \mathcal{F}_t] \\
&= \frac{1 + r_W}{r_W - g} E(FCF_{t+1} z_{t+1} | \mathcal{F}_t) + E(T_C C_{t+1} z_{t+1} | \mathcal{F}_t)
\end{aligned} \tag{13}$$

The second term on the right is

$$E(T_C C_{t+1} z_{t+1} | \mathcal{F}_t) = \frac{T_C r_D D_t}{1 + r_D} = \frac{T_C r_C W_t}{1 + r_D} \frac{L}{1 + L} \tag{14}$$

Using the notation for the adjustment factor in line 5 of Table 2, and combining Eqs.(13) and (14), we get

$$W_t \left(1 - \frac{\lambda L}{1 + L}\right) = \frac{1 + r_W}{r_W - g} E(FCF_{t+1} z_{t+1} | \mathcal{F}_t) \tag{15}$$

Dividing Eq. (15) by Eq. (12)

$$\frac{W_t}{V_t} \left(1 - \frac{\lambda L}{1 + L}\right) = \frac{1 + r_W}{1 + r_A} \frac{r_A - g}{r_W - g} \tag{16}$$

From lines 3 and 4 of Table 2 we have $\frac{W_t}{V_t} = \frac{r_A - g}{r_W - g}$, and Eq. (5) follows.

71.4 Other Models with Fixed Debt or Leverage Ratio

If we modify the assumptions of Modigliani and Miller (1963) to allow debt to grow at a fixed mean rate g , and assume that the cost of debt r_D remains constant, we can formally write a version of Eq. (1) with growth

$$VTS_t = \frac{T_C r_D D_t}{1 + r_D} + \frac{1 + g}{1 + r_D} VTS_t \tag{17}$$

The solution of Eq. (17) represents a Modigliani-Miller formula with growth

$$VTS_t = \frac{T_C r_D D_t}{r_D - g} \tag{18}$$

The formula in Eq. (18) is favored by Myers (1974), Sick (1990), Luehrman (1997), Fieten et al. (2005), and Damodaran (2006). It rests on the heuristic that the uncertainty of tax savings of debt is the same as the uncertainty of debt. There is, however, a counter-argument to this heuristic, which says that the two risks are not the same because default brings the tax savings of debt to zero, but debt frequently has positive recovery rate at default. The Miles-Ezzell formula delivers a refinement of the valuation technique in Eq. (18).

71.4.1 Example

Debt value is $D_t = 250$, constant cost of debt is $r_D = 8\%$, mean growth is $g = 7\%$, and the corporate tax rate is $T_C = 25\%$. Using Eq. (18) we get $VTS_t = 500$.

In the context of the discrete-time stationary model, Fernandez (2004) and Roncaglio and Zanetti (2004) derive a valuation formula of interest tax shields that differs from the Miles-Ezzell formula and states instead that

$$VTS_t = \frac{T_C r_A D_t}{r_A - g} \quad (19)$$

Because the cost of debt r_D is less than the required asset return r_A , valuation produced by Eq. (19) is greater than valuation produced by the Miles-Ezzell formula. Arzac and Glosten [2005] and Cooper and Nyborg (2006) question the derivation of Eq. (19) in both Fernandez (2004) and Roncaglio and Zanetti (2004). Both Arzac and Glosten (2005) and Cooper and Nyborg (2006) contend that Fernandez (2004) assumes constant market leverage ratio but uses Eq. (2) instead of Eq. (9) in his derivation. Fernandez (2005) agrees, and contends that Eq. (19) is suitable for valuation of interest tax shields when the firm maintains constant leverage ratio in book value terms, because then changes in the amount of debt have the same risk as unleveraged free cash flows.

Arzac and Glosten (2005) derive a valuation equation which represents an additive adjustment to the Modigliani-Miller formula. From line 0 of Table 1

$$T_C C_{t+1} = \frac{T_C}{1 - T_C} (FCF_{t+1} - ECF_{t+1}^* + D_{t+1} - D_t) \quad (20)$$

Therefore

$$\begin{aligned} VTS_t &= \sum_{s \geq t} E(T_C C_{s+1} z_{s+1} | \mathcal{F}_t) \\ &= \frac{T_C}{1 - T_C} \sum_{s \geq t} E[FCF_{s+1} z_{s+1} - ECF_{s+1}^* z_{s+1} + (D_{s+1} - D_s) z_{s+1} | \mathcal{F}_t] \quad (21) \\ &= \frac{T_C}{1 - T_C} (V_t - S_t) + \frac{T_C}{1 - T_C} \sum_{s \geq t} E[(D_{s+1} - D_s) z_{s+1} | \mathcal{F}_t] \end{aligned}$$

Taking into account that $V_t - S_t = D_t - VTS_t$, and solving Eq. (21) for VTS_t , we get the formula

$$VTS_t = T_C D_t + T_C \sum_{s \geq t} E[(D_{s+1} - D_s) z_{s+1} | \mathcal{F}_t] \quad (22)$$

Applying Eq. (22) to the Miles-Ezzell framework produces the Miles-Ezzell formula

$$\begin{aligned}
 VTS_t &= T_C D_t + T_C D_t \sum_{s \geq t} \left[\left(\frac{1+g}{1+r_A} \right)^{s-t+1} - \frac{1}{1+r_D} \left(\frac{1+g}{1+r_A} \right)^{s-t} \right] \\
 &= \frac{T_C r_D D_t}{r_A - g}
 \end{aligned} \tag{23}$$

Arzac and Glosten (2005) show that Roncaglio and Zanetti (2004) get Eq. (19) when they disregard, in Eq. (22), the lack of systematic risk between debt at time, t represented by D_t , and the pricing kernel at time $t + 1$, represented by z_{t+1}

$$\begin{aligned}
 VTS_t &= T_C D_t + T_C D_t \sum_{s \geq t} \left[\left(\frac{1+g}{1+r_A} \right)^{s-t+1} - \frac{1}{1+r_A} \left(\frac{1+g}{1+r_A} \right)^{s-t} \right] \\
 &= \frac{T_C r_A D_t}{r_A - g}
 \end{aligned} \tag{24}$$

71.4.2 Example

Debt value is $D_t = 250$, required asset return is $r_A = 12\%$, mean growth is $g = 7\%$, and the corporate tax rate is $T_C = 25\%$. Using Eq. (19) we get $VTS_t = 150$.

Liu (2009) questions the valuation framework of Modigliani and Miller (1963), arguing that because Eq. (2) is invalid when the debt is interest-free, or when the firm has infinite leverage ratio L_t , the Modigliani-Miller formula does not value interest tax shields correctly anywhere in the Modigliani-Miller framework. Instead, Liu (2009) proposes a valuation method, that he calls the slicing approach, of decomposing the Modigliani-Miller formula into earned and unearned parts, with the earned part being $\frac{r_{D_t}}{R_t} T_C D_t$, where R_t is return on invested capital. Qi (2011) delivers a critique of the arguments in Liu (2009), contending that the issue of correctness of Eq. (2) in the Modigliani-Miller framework is resolved by replacing Eq. (2) with Eq. (3), that the special case of infinite leverage does not invalidate Eq. (2), and that the slicing approach is not consistent with prevailing models of the relationship between risk and return.

71.5 Models with Dynamic Debt Policies

The general problem with valuing interest tax shields is that the risk profile of tax shields is different from the risk profile of either debt or equity. Interest tax shields are less variable than equity cash flows, and expire when debt may still have positive recovery rate at default. In addition, fixed leverage ratio is not a very realistic

assumption: In bad times the firm would have to prepay debt and issue new stock at a low price.

The fixed market leverage ratio assumption requires a financing strategy that makes the value of debt a constant multiple of unleveraged free cash flow FCF_t . Indeed, from lines 0 and 0 of Table 2

$$D_t = \frac{1+g}{r_W - g} \frac{L}{1+L} FCF_t \quad (25)$$

Grinblatt and Liu (2008) present a general, continuous-time valuation framework for interest tax shields when unleveraged free cash flow is a general Ito diffusion process, and financing strategy makes the value of debt a linear combination of real powers of unleveraged free cash flow, such as $D_t = d_0 + d_{\frac{1}{3}}FCF_t^{\frac{1}{3}} + d_{\frac{1}{2}}FCF_t^{\frac{1}{2}} + d_1FCF_t + d_2FCF_t^2$. Apart from simple special cases, valuation of interest tax shields requires a numerical solution of a system of ordinary differential equations. The paper does not explicitly model the loss of interest tax shields upon default.

Couch et al. (2012) use barrier option methodology to model explicitly the loss of interest tax shields associated with default, and to allow more flexible financing strategies than the constant rebalancing of debt to maintain a fixed leverage ratio. Recall that the stationary discrete-time model in Sect. 71.3.3 specifies an unleveraged free cash flow that fluctuates around a growing level $FCF_{t+1} = FCF_t(1 + g + \varepsilon_{t+1})$, with noise ε_t drawn from a fixed distribution, and $E(\varepsilon_{t+1} | \mathcal{F}_t) = 0$. The corresponding continuous-time process that they use for EBIT X_t is an Ito diffusion with constant coefficients, where B_t is a standard Brownian motion

$$dX_t = X_t(gdt + \sigma dB_t) \quad (26)$$

As before, expected EBIT grows at a constant rate g , but now there is a second constant parameter σ , which captures the volatility of EBIT, and is commonly considered a measure of business risk.

As in Fischer et al. (1989) and Goldstein et al. (2001), this continuous-time framework can capture an asymmetric financing strategy and explicitly model the possibility of loss of interest tax shields when the firm defaults. Specifically, the firm issues debt with coupon or interest expense C , but does not maintain a fixed leverage ratio. Financing policy is asymmetric: The firm increases the amount of debt at random times when EBIT X_t increases by a specified growth factor, but does not reduce the amount of debt when EBIT declines. Instead, the debt carries an interest coverage ratio covenant, setting a minimum coverage ratio δ of EBIT X_t to coupon C . In other words, leverage ratio fluctuates between refinancing times, and the firm defaults if and when EBIT X_t first falls to the level $H = \delta C$ before the next refinancing.

Denote by X_0 the level of EBIT X_t at the time when the debt was issued, and by ϕ a constant greater than 1. The firm's financing strategy is to refinance the debt if and when EBIT X_t first reaches the level $K = \phi X_0$ before defaulting, at which time the firm repays the outstanding debt with coupon C and issues new debt with coupon ϕC . After refinancing, the minimum coverage ratio is again δ , so that the firm defaults on the new debt if and when EBIT X_t first falls to the level ϕH before reaching the level ϕK , and refinances if and when EBIT first hits the level ϕK before reaching the level ϕH . The refinancing process repeats itself if and when EBIT reaches the levels $\phi^2 K$, $\phi^3 K$, etc., with the firm defaulting if and when EBIT reaches one of the levels $\phi^2 H$, $\phi^3 H$, etc.

Let $H \leq X \leq K$ and denote by $P_{HK}(X; H, K)$ the price of a double-barrier option that pays \$1 when EBIT X_t , starting from X , first hits the barrier H before hitting the barrier K . Similarly, denote by $Q_{HK}(X; H, K)$ the price of a double-barrier option that pays \$1 when EBIT X_t , starting from X , first hits the barrier K before hitting the barrier H . Then the value $VTSX_0$ of interest tax shields when EBIT is X_0 satisfies the recursive equation

$$VTS(X_0) = \frac{T_C C}{r_F} [1 - P(X_0; H, K) - Q(X_0; H, K)] + \phi VTS(X_0) Q(X_0; H, K) \quad (27)$$

Solving Eq. (27) for $VTSX_0$, we get

$$VTS(X_0) = \frac{T_C C}{r_F} \frac{1 - P(X_0; H, K) - Q(X_0; H, K)}{1 - \phi Q(X_0; H, K)} \quad (28)$$

Equipped with the value of $VTSX_0$, we can calculate the value of interest tax shields for any level X of EBIT X_t such that $H \leq X \leq K$ by writing

$$VTS(X) = \frac{T_C C}{r_F} [1 - P(X; H, K) - Q(X; H, K)] + \phi VTS(X_0) Q(X; H, K) \quad (29)$$

Combining Eqs.(28) and (29) and using shorthand notation for $PX; H, K$ and $QX; H, K$, we get a valuation formula for any $H \leq X \leq K$

$$VTS(X) = \frac{T_C C}{r_F} \left[1 - P - Q + (1 - P_0 - Q_0) \frac{\phi Q}{1 - \phi Q_0} \right] \quad (30)$$

We can think of the term $1 - P - Q + (1 - P_0 - Q_0) \frac{\phi Q}{1 - \phi Q_0}$ in Eq. (30) as a risk adjustment factor for either the amount of interest tax shield $T_C C$ or the corporate tax shield T_C .

The valuation and properties of barrier options have been studied by Rubinstein and Reiner (1991) and Derman and Kani (1996,1997). Theoretical prices of the double-barrier options used here have also been derived by Gerber and Shiu (1994)

using martingale methods, and by Goldstein et al. (2001) using partial differential equations. For any $H \leq X \leq K$ we have

$$\begin{aligned} P(X; H, K) &= \frac{\left(\frac{H}{K}\right)^{b_m} \left(\frac{H}{X}\right)^{b_p} - \left(\frac{H}{K}\right)^{b_p} \left(\frac{H}{X}\right)^{b_m}}{\left(\frac{H}{K}\right)^{b_m} - \left(\frac{H}{K}\right)^{b_p}} \\ Q(X; H, K) &= \frac{\left(\frac{H}{X}\right)^{b_m} - \left(\frac{H}{X}\right)^{b_p}}{\left(\frac{H}{K}\right)^{b_m} - \left(\frac{H}{K}\right)^{b_p}} \end{aligned} \quad (31)$$

where

$$\begin{aligned} a &= \frac{g - (r_A - r_F)}{\sigma^2} - \frac{1}{2} \\ \text{and } b_m &= a - \sqrt{a^2 + \frac{2r_F}{\sigma^2}} \end{aligned} \quad (32)$$

71.5.1 Example

The firm's EBIT is $X = 100$, debt was originally issued at EBIT $X_0 = 60$, with coupon $C = 20$, and minimum interest coverage ratio $\delta = 1.5$. First refinancing will occur at $K = 120$ and, thereafter, at multiples $\phi = 2$ of the previous level of refinancing. Mean growth rate of EBIT is $g = 7\%$, volatility of EBIT is $\sigma = 20\%$, risk-free interest rate is $r_F = 6\%$, required asset return is $r_A = 12\%$, and corporate tax rate is $T_C = 25\%$.

Then the default barrier is $H = 30$, the refinancing barrier is $K = 120$, and, from Eq. (32), the risk exponents are $b_p = 1.5000$ and $b_m = -2.0000$. Therefore, from Eq. (31), the prices of the double-barrier options at the time of debt issue were $PX_0; H, K = 0.3248$ and $QX_0; H, K = 0.2297$. From Eq. (28), we get the value of interest tax shields at the time of debt issue as $VTSX_0 = 68.67$.

In contrast to the models in Sects. 71.3 and 71.4, the value of interest tax shields in this framework depends on both business risk and the level of EBIT. Table 3 presents sensitivity analysis of this relationship, changing one parameter at a time while keeping the other parameters constant. Thus, the values in Table 3 assume that the firm may change business risk σ without changing systematic risk as reflected in the required asset return r_A . If both business risk and systematic risk change, we have to use Eq. (30) to compute the new value of interest tax shields. It is also important to keep in mind that the tax benefits in Table 3 are gross of deadweight default cost, and that subtracting the value of deadweight default cost reduces those tax benefits.

For a comparison, Examples 71.3.2, 71.3.4, 71.4.1, and 71.4.2 use the data $D_t = 250$, $r_D = 8\%$, $r_A = 12\%$, $g = 7\%$, $C_t = 20$, and $T_C = 25\%$, and produce the

Table 3 Value of interest tax shields for a range of EBIT and business risk

X	σ				
	0.10	0.15	0.20	0.25	0.30
80	110.52	105.73	92.12	77.20	63.91
100	139.02	131.63	114.06	95.31	78.77
120	180.41	162.70	137.33	113.11	92.68
140	237.84	200.41	163.03	131.50	106.40
160	314.13	245.59	191.66	150.90	120.30

Table 4 Comparison of valuations of interest tax shields

1	2	3	4	5	6
Modigliani-Miller	Miles-Ezzell	Modified Miles-Ezzell	Modigliani-Miller with Growth	Fernandez	Couch-Dothan-Wu
62.50	103.70	100.00	500.00	150.00	82.73

valuations presented in columns 1–5 of Table 4. The calculation for column 6 is shown below.

To calculate the value in column 6, we need to compute EBIT X for which the market value of debt is $D_t = 250$ that was used in all the previous examples. Denote by F the face value of debt, by θ the deadweight default cost as percentage of enterprise value at default, and set $\sigma = 25\%$ and $\theta = 30\%$. The choice of θ is consistent with empirical findings for liquidation cost reported by Gilson (1997) to average 36.5%, and by Alderson and Betker (1995) to average 45.3%. Hackbarth et al. (2006) use 40.0%, computed as an approximate average of those two numbers. Our choice of a smaller value $\theta = 30\%$ reflects the fact that it represents a default cost rather than liquidation cost. Enterprise value at default is calculated as $WH = 225.00$.

Similar to the analysis of Fischer et al. (1989) and Goldstein et al. (2001) we get a valuation formula for debt

$$D(X_0) = \frac{C}{r_F} [1 - P(X_0; H, K) - Q(X_0; H, K)] + \min [(1 - \theta)W(H), F] P(X_0; H, K) + FQ(X_0; H, K) \tag{33}$$

Assume that debt was originally issued at par, then $DX_0 = F$. Solving Eq. (33) and $DX_0 = F$ for face value F , we get $F = 235.30$. Using this face value, it is easy to calculate the cost of debt at issue and recovery rate at default as percentage of face value

$$\begin{aligned} \text{Cost of debt at issue } r_{0,D} &= \frac{C}{F} = 8.50\% \\ \text{Recovery rate at default} &= \frac{(1 - \theta)W(H)}{F} = 66.94\% \end{aligned} \tag{34}$$

It is interesting to note that this recovery rate at default is within one standard deviation of the mean recovery rate for U.S. nonfinancial corporations over the past 20 years, reported by Covitz and Han (2004) as $40 \pm 28\%$. Given F , we can rewrite Eq. (33) with X instead of X_0 , set $DX = D_t$, and solve the resulting equation for X ,

yielding $X = 85.97$. As constructed, when $X = 85.97$, the market value of debt is $DX = 250.00$, and cost of debt is $r_{D,t} = \frac{C}{D_t} = 8.00\%$. Then, with $\sigma = 25\%$ and $X = 85.97$, the value of interest tax shields, computed from Eq. (30), is $VTS_t = 82.73$. This value is greater than the Modigliani-Miller value because of positive growth, and less than the Miles-Ezzell (and modified Miles-Ezzell) value because of the asymmetric and infrequent rebalancing of debt, and the explicit modeling of the loss of interest tax shields if and when the firm defaults.

71.6 Conclusion

Financing effects are an important part of valuation, and the adjusted present value method (APV) helps make those effects more visible. To perform APV valuation, we need to assess the value of interest tax shields.

The finance literature offers several popular methods of valuation of interest tax shields. The methods are based on different sets of assumptions and/or heuristics, often limiting the firm's financing strategies, and produce somewhat different results. In this chapter, I reviewed and illustrated several such methods, and commented on their validity and applicability.

In addition, researchers are developing new, dynamic models of valuation of interest tax shields that allow more flexible financing strategies. The chapter reported on such new work, and presented a summary that compares the valuation outcomes of the various methods.

References

- Alderson, M.J., and B.L. Betker. 1995. Liquidation costs and capital structure. *Journal of Financial Economics* 39: 45–69.
- Arzac, E.R., and L.R. Glosten. 2005. A reconsideration of tax shield valuation. *European Financial Management* 11: 453–461.
- Booth, L. 2002. Finding value where none exists: Pitfalls in using adjusted present value. *Journal of Applied Corporate Finance* 15: 95–104.
- Brealey, R., S. Myers, and F. Allen. 2011. *Principles of corporate finance*. 10th ed. New York: McGraw-Hill/Irwin.
- Cooper, I., and K.G. Nyborg. 2006. The value of tax shields IS equal to the present value of tax shields. *Journal of Financial Economics* 81: 215–225.
- Couch, R., M. Dothan, and W. Wu. 2012. Interest tax shields: A barrier options approach. *Review of Quantitative Finance and Accounting* 39: 123–146.
- Covitz, D., and S.Han. 2004. *An empirical analysis of bond recovery rates: Exploring a structural view of default*. Working paper, Federal Reserve Board.
- Damodaran, A. 2006. *Damodaran on valuation*. 2nd ed. Hoboken: Wiley.
- Derman, E., and I. Kani. 1996. The ins and outs of barrier options: Part 1. *Derivatives Quarterly* 3: 85.
- . 1997. The ins and outs of barrier options: Part 2. *Derivatives Quarterly* 4: 52.
- Farber, A., R. Gillet, and A. Szafarz. 2006. A general formula for the WACC. *International Journal of Business* 11: 211–218.
- Fernandez, P. 2004. The value of tax shield is not equal to the present value of tax shields. *Journal of Financial Economics* 73: 145–165.

- . 2005. *The value of tax shields depends only on the net increases of debt*. Working paper, University of Navarra.
- Fieten, P., L. Kruschwitz, J. Laitenberger, A. Löffler, J. Tham, I. Velez-Pareja, and N. Wonder. 2005. Comment on "The value of tax shields is not equal to the present value of tax shields". *Quarterly Review of Economics and Finance* 45: 184–187.
- Fischer, E.O., R. Heinkel, and J. Zechner. 1989. Dynamic capital structure choice: Theory and tests. *Journal of Finance* 44: 19–40.
- Gerber, H.U., and E.S.W. Shiu. 1994. Martingale approach to pricing perpetual American options. *ASTIN Bulletin (Casualty Actuarial Society)* 24: 195–220.
- Gilson, S.C. 1997. Transactions costs and capital structure choice: Evidence from financially distressed firms. *Journal of Finance* 52: 161–196.
- Goldstein, R., N. Ju, and H. Leland. 2001. An EBIT-based model of dynamic capital structure. *Journal of Business* 74: 483–512.
- Graham, J.R. 2000. How big are the tax benefits of debt? *Journal of Finance* 55: 1901–1941.
- Grinblatt, M., and J. Liu. 2008. Debt policy, corporate taxes, and discount rates. *Journal of Economic Theory* 141: 225–254.
- Hackbarth, D., J. Miao, and E. Morellec. 2006. Capital structure, credit risk, and macroeconomic conditions. *Journal of Financial Economics* 82: 519–550.
- Harris, R.S., and J.J. Pringle. 1985. Risk-adjusted discount rates: Extensions from the average risk case. *Journal of Financial Research* 8: 237–244.
- Kaplan, S.N., and P. Strömberg. 2009. Leveraged buyouts and private equity. *Journal of Economic Perspectives* 23: 121–146.
- Kemsley, D., and D. Nissim. 2002. Valuation of the debt tax shield. *Journal of Finance* 57: 2045–2073.
- Koller, T., M. Goedhart, and D. Wessels. 2010. *Valuation: Measuring and managing the value of companies*. 5th ed. Hoboken: Wiley.
- Lewellen, W.G., and D.R. Emery. 1986. Corporate debt management and the value of the firm. *Journal of Financial and Quantitative Analysis* 21: 415–426.
- Liu, Y.-C. 2009. The slicing approach to valuing tax shields. *Journal of Banking and Finance* 33: 1069–1078.
- Luehrman, T.A. 1997. Using APV: A better tool for valuing operations. *Harvard Business Review* 75: 145–154.
- Miles, J.A., and J.R. Ezzell. 1980. The weighted average cost of capital, perfect capital markets, and project life: A clarification. *Journal of Financial and Quantitative Analysis* 15: 719–730.
- . 1985. Reformulation of tax shield valuation: A note. *Journal of Finance* 40: 1485–1492.
- Modigliani, F., and M.H. Miller. 1958. The cost of capital, corporation finance, and the theory of investment. *American Economic Review* 48: 261–297.
- . 1963. Corporate income taxes and the cost of capital: A correction. *American Economic Review* 53: 433–443.
- Myers, S.C. 1974. Interactions of corporate financing and investment decisions – Implications for capital budgeting. *Journal of Finance* 29: 1–25.
- Qi, H. 2011. Value and capacity of tax shields: An analysis of the slicing approach. *Journal of Banking and Finance* 35: 166–173.
- Roncaglio, F., and L. Zanetti, 2004. *On the equivalence between the APV and the WACC approach in a growing levered firm*. Working paper, Bocconi University.
- Ruback, R.S. 1995. *A note on capital cash flow valuation*. Teaching note 9-295-069, Harvard Business School.
- . 2002. Capital cash flows: A simple approach to valuing risky cash flows. *Financial Management* 31: 85–103.
- Rubinstein, M., and E. Reiner. 1991. Unscrambling the binary code. *Risk* 4: 75–83.
- Sick, G.S. 1990. Tax-adjusted discount rates. *Management Science* 36: 1432–1450.
- Taggart, R.A. 1991. Consistent valuation and cost of capital expressions with corporate and personal taxes. *Financial Management* 20: 8–20.



Savita A. Sahay

Abstract

This chapter aims to study the current rules regarding preparation of cash flow statements, as proscribed by Financial Accounting Standards Board (FASB) and pinpoint the problems with the current rules.

The problems with the existing rules under SFAS-95 can be classified into four broad categories: problems arising from the current three-way classification of cash flows, problems due to the method chosen for preparation of cash flow statements, problems due to the ambiguities in the current rules and problems due to the loopholes in the rules that allow manipulation of cash flows. Several examples of each inconsistency and ambiguity are provided using Johnson & Johnson (JNJ) cash flow statements for 2011.

We use an illustrative method to highlight the need for changes in the existing rules regarding cash flow statements. By showing that under the current rules, several cash flows get misclassified, leading to unclear and confusing financial reports, we hope to make the readers of the financial statements aware that the amounts reported as cash flow from operating, investing, and financing activities do not reflect all of the cash consequences of these activities and that the reported amounts should not be used in decision models without adjustment. We also hope that the FASB will be persuaded to amend SFAS-95 to eliminate the ambiguities and conform more closely to the finance literature, so that the information in the cash flow statement becomes more useful to report users.

Since a primary objective of financial reporting is to provide information to help accounting users assess the amount and timing of future cash flow and earnings, removing some of the ambiguities and other improvements suggested in the paper will not only make these statements more useful in credit and

S. A. Sahay (✉)

Department of Accounting and Information Systems, Rutgers Business School, Rutgers University, Piscataway, NJ, USA

e-mail: savita.sahay@business.rutgers.edu

investment decisions, it will also lead to an increased usefulness of financial ratios which currently use information from the cash flow statements, such as operating cash flow, in either the numerator or denominator.

Keywords

Cash classification ambiguities · Cash flow statement · Existing rules · Financial ratios · Manipulation · SFAS-95

72.1 Introduction

A flood of literature in accounting on earnings management indicates the obsession of auditors, accountants, investors and analysts with reported earnings number.¹ However, a number of recent corporate scandals are bringing home the realization that corporate management can manipulate earnings in various ways. Furthermore, accrual accounting fails to provide adequate information about the liquidity of the firm and its long-term solvency. As a result, an increasing number of investors and analysts have turned their focus towards the Statement of cash flows (SCF).

The statement of cash flow plays an important role in fundamental securities analysis.² It is intended to help predict the firm's ability to sustain (and increase) cash from current operations. As noted in the statement of Financial Accounting standards (SFAS) No. 95, a SCF "provides useful information about an entity's activities in generating cash through operations to repay debt, distribute dividends, or reinvest to maintain or expand operating capacity." For example, comparing operating cash flow with net income is necessary to assess the reasons for the differences between net income and cash flows. Similarly, Investing and financial cash flows can reveal how a firm's asset mix and financing mix is changing. Analysis of cash flow statement can also provide useful insights into the company's ability to generate future cash flows, pay dividends and meet other future obligations, reveal trends in cash flow components and cash consequences of investing and financing decisions.³ The statement can also convey many crucial management decisions regarding financial policy (leverage), dividend policy and investment policies for future growth. Moreover, Many Cash flow ratios, valuable in evaluating a firm's short term as well as long-term liquidity, use cash flow from operations and other information in the SCF.⁴ Thus, it is necessary that this statement is as transparent and user friendly as possible.

Given the vital importance of cash flow statements, it is surprising that while a plethora of authoritative guidance surrounds the calculation and presentation of

¹See Finger (1994).

²See for example, Barth et al. (1999 and 2001).

³See D fisher, "Cash doesn't lie," Forbes Magazine (April 12, 2010), pp. 52–55.

⁴Ali, A. 1994.

earnings, existing FASB rules say so little on this important piece of information. Only SFAS 95 specifically addresses the cash flow statement, and only 15 paragraphs within SFAS 95 discuss the appropriate categorization of cash expenditures within the cash flow statement. Moreover, the current format of cash flow statement does not provide either the cash flow from assets or the total of cash flow to creditors and stockholders, a number most useful to analysts, investors and finance experts.⁵

This chapter aims to study the current rules regarding preparation of cash flow statements, as proscribed by Financial Accounting Standards Board (FASB) and pinpoint the problems with the current rules. Some of these problems are due to the arbitrariness of the classifications by FASB. Some others are due to the method chosen for preparation of cash flow statements while others stem from loopholes in the rules that allow manipulation of cash flows in much the same way as earnings manipulation.

The chapter hopes to highlight that the three-way classification under SFAS-95 results in reporting similar cash flows differently and different cash flows similarly, just the opposite of the desired characteristic of financial reporting. Several examples of each inconsistency and ambiguity are provided using Johnson & Johnson (JNJ) cash flow statements for 2011.

By highlighting these inconsistencies and ambiguities of SFAS-95, we hope that users of the statements will be more aware that the amounts reported as cash flow from operating, investing, and financing activities do not reflect all of the cash consequences of these activities and that the reported amounts should be used cautiously in decision models. It is also hoped that the FASB will be persuaded to amend SFAS-95 to eliminate the ambiguities and more closely conform the classification rules to the finance literature, so that the information in the cash flow statement more closely conforms to the decision models and therefore becomes more useful to report users.

72.2 FASB Rules for Cash Flow Statements

FASB Statement No. 95 (SFAS-95)⁶ requires firms to provide a cash flow statement for each period for which results of operations are provided. The cash flow statement explains the change in cash and cash equivalents, and classifies cash receipts and cash payments as resulting from operating, investing, or financing activities. In addition, firms with significant foreign operations are required to report a fourth category called the effect of foreign exchange rate changes on cash, which accumulates the effects of changes in exchange rates on foreign currency transactions.

⁵See Bowen et al. (1986), Dechow (1994) and Dechow et al. (1998).

⁶See FASB (1995).

72.2.1 Operating Activities

In SFAS-95, the FASB defines operating activities as a residual category to include all transactions and events other than investing and financing activities. It notes that operating activities generally relate to producing and delivering goods and providing services, and that the cash flows from operating activities are generally the cash effects of transactions and other events that enter into the determination of net income.

Specifically, operating cash inflows include customer collections from the sales of goods and services, interest and dividend collections on debt and equity securities of other entities, and all other receipts not defined as investing or financing inflows, such as supplier refunds, collections on lawsuits, and most insurance proceeds. Operating cash outflows include interest payments (unless capitalized), payments for inventories, payments to employees, payments to suppliers of other goods and services, payments to governments for taxes, duties, fines, and other fees, and all other payments not defined as investing or financing outflows, such as customer refunds, payments under lawsuits, and charitable contributions.

72.2.2 Investing Activities

SFAS-95 defines investing activities to include making and collecting loans and acquiring and disposing of debt or equity instruments and plant assets.

Thus, Investing cash inflows include receipts from collecting loans, receipts from sales of debt or equity securities of other entities, and receipts from sales of plant assets. Note that under SFAS-95, interest and dividend collections are classified as operating inflows, not investing inflows. Investing cash outflows include payments to make or acquire loans, payments to acquire debt or equity securities of other entities, and payments to acquire plant assets.

72.2.3 Financing Activities

SFAS-95 defines financing activities to include obtaining resources from owners and providing them with a return on their investment; borrowing money and repaying amounts borrowed, or otherwise settling the obligation; and obtaining and repaying other resources obtained from creditors on long-term credit.

Under SFAS-95, financing cash inflows include proceeds from issuing debt or equity securities and from other short- or long-term borrowings. Financing cash outflows include dividend payments, outlays to reacquire or retire equity securities, and repayments of amounts borrowed.

Note that under SFAS-95, interest payments are classified as operating outflows, not financing outflows. So while interest and dividends received is an operating cash inflow, interest payments and dividends paid are financing cash outflow.

The rationale offered in SFAS-95 for the three-way distinction among operating, investing, and financing cash flows is that investors and creditors consider the

relationships among certain cash flows to be important to their analysis of financial performance, and that classifying cash flows as operating, investing, and financing facilitates the evaluation of significant relationships within and among these basic activities. This classification, the FASB notes “links cash flows that are often perceived to be related, such as cash proceeds from borrowing transactions and cash repayments of borrowings . . . [and results in a cash flow]. . . statement which reflects the cash flow impact of each of the major activities of the enterprise.” In general, operating activities involve income statement items, investing activities involve cash flows resulting from changes in investment and long term assets and financing activities involve cash flow resulting from changes in long term liability and stockholders’ equity.

The finance literature notes that investing and financing decisions are interrelated and that both are related to operating decisions. Nevertheless, the finance literature posits that, for many purposes, it is useful to distinguish between investing decisions and financing decisions and to separate investing and financing decisions from operating decisions.

Comparing the three-way classification of SFAS-95 to the finance literature emanates from the decision usefulness orientation of financial reporting. Because a primary objective of financial reporting is to provide information that is useful in credit and investment decisions, cash flow statements should provide information needed for credit and investment decision models. Many of these decision models are specified in the finance literature and use financial ratios. For example, analysts, researchers and textbooks calculate certain financial ratios that use operating cash flow in either the numerator or denominator. The usefulness of these ratios as well as cash flow information generally should be enhanced by classifying cash flows in the cash flow statement as they are classified in the finance literature.

Although the three-way classification of SFAS-95 is loosely based on the finance literature, certain modifications result in inconsistent or ambiguous classifications of certain cash flows. Accordingly, reported cash flow from operating activities under SFAS-95 frequently includes cash flows from investing and financing activities. Similarly, reported cash flows from investing activities and financing activities frequently exclude certain cash flows properly attributable to them.

The aim of this chapter is to emphasize that the operating cash flows should reflect the cash effects of items included in calculating operating income for the same period. Thus, the operating section of the SCF should provide a cash basis measure of the company’s performance in much the same way that income statement provides an accrual basis measure.

72.3 Statement of Cash Flows: Methods

The general format of the statement lists the three activities described above. The operating section always appears first, followed by the investing activities and then the financing activities. The net increase or decrease in cash reported should reconcile the beginning and ending cash balances in the comparative balance sheets.

There is some flexibility in summarizing the sources and uses of funds in a cash flow statement. Although the type of information presented is similar, the exact order and level of aggregation may differ from one preparer to the next. As we will see in the cash flow statement of J&J, a large amount of aggregation is used under “Other Assets” and “other Liabilities” categories in the cash flow statement, making it very hard to understand and analyze these numbers.

Both FASB and IASB standards allow either the direct or the indirect method to determine cash flows, but both encourage DM presentation.⁷ However, a large majority of firms currently disclose cash flows using the indirect method. Since the firms are allowed to choose either method for a cash flow statement, a deeper look into the properties of firms that self select themselves into Indirect Method will help standard setters in understanding the costs and benefits firms face when making a choice.

The indirect method, also called the reconciliation method, starts with net income and adjusts it for items that affected net income but did not affect cash. To calculate net cash flow from operating activities, the companies add back all non-cash charges to net income and deduct all non-cash credits.

For consistency and comparability, most companies prefer to use indirect method for cash flow statements, although FASB encourages the use of the direct method.

The direct method, also called the income statement method, reports net cash flow from operating activities as the difference between cash receipts and cash payments from operations. In other words, the direct method simply deducts operating cash payments from operating cash receipts.⁸ Cash collections typically include cash collected from customers, from interest and dividends or any other sources. Cash payments include cash paid to suppliers for goods, employees for services, to creditors for interest and to government for taxes.

Since the reporting of investing and financing activities is identical under the two methods, we will focus more on operating activities, since the reporting of cash from operations is quite different under the two methods.

72.3.1 Johnson and Johnson Company: An Example

To facilitate discussion of our objections to the current format of the cash flow statement, we use the example of Johnson and Johnson (J&J) for the most recent year.

It is important to note that the preparation of the cash flows statement is different from the other three basic financial statements – It is not prepared from the adjusted trial balance since the statement requires information concerning changes in account

⁷“Is the direct method for the statement of cash flow preferable to the indirect method, and if so why?” Joint research initiative by FASB and IASB, 2005.

⁸A CFA institute monograph, 2005 on financial reporting lists direct method as one of the 12 significant reforms needed to improve financial reporting.

balances over a period of time. The information to prepare this statement comes from comparative balance sheets, current income statement (Since the net income must be adjusted for the effects of accruals to determine cash flows) and selected transaction data to determine how the company used cash during the period.

Tables 1 and 2 show comparative balance sheets and the income statement for J&J:

A review of the income statement reveals information that would be used in preparing the SCF. There are two important facts to be noted about the income statement: One is that the company does not report any depreciation on the statement and second that interest revenue is included in operating income. In most cases, Companies place interest received in the investing section but interest paid is shown in the operating section. Similarly, most firms place depreciation and gain or loss from sale of assets in the operating section but cash paid for operating assets is shown in the investing section. As shown below in the SCF, J&J shows both interest received and interest paid as operating activities and shows all asset related transactions, including depreciation, in the investing section.

Table 3 shows the statement of cash flows for J&J as submitted to the SEC. Note that like most other submitted statement,⁹ J&J also uses the **indirect method**.

72.3.1.1 Cash Flow from Operating Activities

In the above statement, the company started with the net income, as reported in the income statement, and adjusted it for all:

1. Non cash revenues and expenses (For example, depreciation expense)
2. Increases (decreases) in the balances of all operating assets are subtracted (added): For example, when accounts receivable increase during the year, it implies that cash receipts were lower than revenues. J&J's accounts receivable increased by \$207,000 during the year. This amount will be deducted from the net income to arrive at the amount of cash from operations. Similarly, increase in inventories, or any other operating asset, represents an operating use of cash, even though it is not an expense. This will also be deducted from net income to arrive at the cash flow from operations.
3. Increases (decreases) in the balances of all operating liability accounts are added (subtracted). For example, when accounts payable increase during the year, it means that company paid less cash than the expenses incurred. J&J's accounts payable increased by \$20,000 during the year. This amount will be added back to net income to arrive at the amount of cash from operations. A similar adjustment is made for the increase in deferred taxes or any other expenses payable, which keeps cash in the business.

⁹Accounting Trends and Techniques, 2010, reports that out of its 500 surveyed companies, 495(99%) used the indirect method.

Table 1 Balance sheet (J&J)

	2011	2010	Change
Cash/equivalents	19,355,000.00	15,810,000.00	3,545,000.00
Marketable securities	8,303,000.00	3,615,000.00	4,688,000.00
Accounts receivables – gros	10,114,000.00	9,979,000.00	135,000.00
Doubtful account	– 340,000.00	– 333,000.00	– 7,000.00
Raw materials	1,073,000.00	1,144,000.00	– 71,000.00
Goods in process	1,460,000.00	1,395,000.00	65,000.00
Finished goods	2,845,000.00	2,641,000.00	204,000.00
Deferred taxes	2,224,000.00	2,793,000.00	– 569,000.00
Prepaid and others	2,273,000.00	2,497,000.00	224,000.00
Total current assets	47,307,000.00	39,541,000.00	7,766,000.00
Land improvement	738,000.00	714,000.00	24,000.00
Buildings	9,079,000.00	8,863,000.00	216,000.00
Machinery and equipment	18,032,000.00	17,153,000.00	879,000.00
Construction	2,577,000.00	2,521,000.00	56,000.00
Depreciation	– 15,873,000.00	– 14,492,000.00	– 1,381,000.00
Goodwill	15,294,000.00	14,862,000.00	432,000.00
Intangible – gross	22,225,000.00	21,180,000.00	1,045,000.00
Amortization of intangible	– 5,509,000.00	– 4,857,000.00	– 652,000.00
Deferred taxes	5,096,000.00	5,507,000.00	– 411,000.00
Other assets	3,942,000.00	3,690,000.00	252,000.00
Total assets	102,908,000.00	94,682,000.00	8,226,000.00
Loans and notes payable	7,617,000.00	6,318,000.00	1,299,000.00
Accounts payable	5,623,000.00	5,541,000.00	82,000.00
Accrued liabilities	4,100,000.00	4,625,000.00	– 525,000.00
Accrued compensation	2,642,000.00	2,777,000.00	– 135,000.00
Accrued rebates	2,512,000.00	2,028,000.00	484,000.00
Income taxes	578,000.00	442,000.00	136,000.00
Total current liabilities	23,072,000.00	21,731,000.00	1,341,000.00
Long-term debt	9,156,000.00	8,223,000.00	933,000.00
Total long term debt	9,156,000.00	8,223,000.00	933,000.00
Deferred taxes	1,447,000.00	1,424,000.00	23,000.00
Employee oblig.	6,087,000.00	6,769,000.00	– 682,000.00
Other liabilities	6,567,000.00	5,947,000.00	620,000.00
Total liabilities	46,329,000.00	44,094,000.00	2,235,000.00
Common stock	3,120,000.00	3,120,000.00	0.00
Retained earnings	77,773,000.00	70,306,000.00	7,467,000.00
Compreh. income	– 3,531,000.00	– 3,058,000.00	– 473,000.00
Treasury stock	– 20,783,000.00	– 19,780,000.00	– 1,003,000.00
Total equity	56,579,000.00	50,588,000.00	5,991,000.00
Total liabilities and shareholders' equity	102,908,000.00	94,682,000.00	8,226,000.00

Table 2 Consolidated statements of earnings (J&J)

	For the year 2011 (in millions)
Sales to customers	\$ 61,587
Cost of products sold	18,792
Gross profit	42,795
Selling, marketing and administrative expenses	19,424
Research and development expense	6,844
Purchased in-process research and development (Note 20)	
Interest income	−107
Interest expense, net of portion capitalized (Note 4)	455
Other (income) expense, net	−768
Restructuring (Note 22)	
Earnings before provision for taxes on income	16,947
Provision for taxes on income (Note 8)	3,613
Net earnings	\$ 13,334

A correct procedure to arrive at the amount of net cash flow from operations is to measure change in each operating account on balance sheet to determine whether any changes in balance sheet accounts caused an increase or decrease in cash. A few important points should be noted here:

1. Changes in operating accounts shown on J&J's SCF do not equal the balance sheet changes. For example, the SCF shows that accounts receivables increased by \$207,000. However, the balance sheet reports an increase of only \$135,000. Similarly, accounts payable increase by \$20,000 as per the SCF but the balance sheet shows an increase of \$82,000 in accounts payable. It is not clear what accounts for these and other discrepancies in operating assets and liabilities. Perhaps the company has made some adjustments and aggregations, but they are not explained.
2. The SCF shows an increase of deferred taxes of \$356,000. A careful look at the balance sheet reveals that deferred taxes are reported thrice: In current assets, in long-term assets and also in long-term liabilities. Moreover, the total of the three is not equal to the amount reported in the SCF.
3. J&J reports an increase in almost all of its long-term assets in the balance sheet, (Land, building, machinery and equipment, construction and goodwill all went up significantly) but the SCF provides only a single number of additions to property, plant and equipment of \$2,384,000, which does not add up to all the increases in asset. Moreover, the SCF reports proceeds from sale of assets in the investing section, but which assets were sold and whether they were sold at a gain or a loss is not revealed. The company also reports significant transactions in investments and in short term debt instruments in the SCF but balance sheet does not even have a category named investments in its asset section. Moreover, the short-term debt does not match the short-term loans payable in the balance sheet. Why is this

Table 3 Indirect method Johnson & Johnson (J&J) company

Statement of cash flow for the year ended December 31, 2011	
Net income/starting line	13,334,000.00
Depreciation	2,939,000.00
Depreciation/depletion	2,939,000.00
Deferred taxes	356,000.00
Other non-cash items	626,000.00
Non-cash items	626,000.00
Accounts receivable	-207,000.00
Inventories	-196,000.00
Other assets	-574,000.00
Payable/accrued	20,000.00
Other liabilities	87,000.00
Changes in working capital	-870,000.00
Cash from operating activities	16,385,000.00
Purchase of fixed assets	-2,384,000.00
Capital expenditures	-2,384,000.00
Acquisition of business	-1,269,000.00
Sale of fixed assets	524,000.00
Sale/maturity of investment	11,101,000.00
Purchase of investments	-15,788,000.00
Other investing cash flow	-38,000.00
Other investing cash flow items, total	-5,470,000.00
Cash from investing activities	-7,854,000.00
Cash dividends paid – common	-5,804,000.00
Total cash dividends paid	-5,804,000.00
Repurchase/retirement of common	-2,797,000.00
Common stock, net	-2,797,000.00
Options exercised	1,226,000.00
Issuance (retirement) of stock, net	-1,571,000.00
Short term debt issued	7,874,000.00
Short term debt reduction	-6,565,000.00
Short term debt, net	1,309,000.00
Long term debt issued	1,118,000.00
Long term debt reduction	-32,000.00
Long term debt, net	1,086,000.00
Issuance (retirement) of debt, net	2,395,000.00
Cash from financing activities	-4,980,000.00
Foreign exchange effects	-6,000.00
Net change in cash	3,545,000.00
Net cash – beginning balance	15,810,000.00
Net cash – ending balance	19,355,000.00
Cash interest paid	491,000.00
Cash taxes paid	2,442,000.00

information not deemed important and why the cash paid for operating assets is not shown in the operating section are major concerns that FASB should address.

72.3.1.2 Cash Flow from Investing Activities

Capital expenditures for long-term assets such as plant, property and equipment, land and Buildings are the primary component of investing cash flow. Capital expenditures may be reported net or gross of proceeds from the sale of these assets. However, trends in gross capital expenditures contain useful insight into management plans. The J&J cash flow statement reports only a net figure for all plant property and equipment, without providing details about changes in their composition.

Other components of cash flow from investing activities include consequences of acquisitions, investments in affiliates and joint ventures. However, segregating operating assets and liabilities acquired in acquisitions may provide useful information about payments made in cash. The J&J cash flow statement reports a capital expenditure of \$2,384,000 for purchase of various fixed assets. Moreover, they report a cash outflow of \$1,269,000 for acquisition of business and another cash outflow of \$5,470,000 as “Other cash outflow.” It is clear that the company engaged in significant acquisition activity during the year. Since most of the company’s long term assets went up during the year, it is impossible to figure out how much of the increase in assets is due to purchase of new assets and how much is due to acquisition of another business and its assets and liabilities. This is probably the reason why the changes in balance sheet accounts do not correspond to changes in the cash flow statement.

72.3.1.3 Cash Flow from Financing Activities

Cash inflows from financing activities include additional borrowing and equity financing and cash outflows include repayment of debt, dividend payments and equity repurchase or retirement. The calculation of equity financing cash flows requires separation of net income, dividend declared and changes in valuation accounts included in equity. The J&J cash flow statement reports a significant cash payment for dividends and repurchase of its common stock. The debt undertaken also increased in both short term and long term categories.

Once the cash flow from operating, investing and financing operations is calculated, the last step is to summarize the net cash flow, which, by definition, must be equal to the change in cash. As indicated at the bottom of the SCF, net increase in cash of \$3,545,000 is the same as shown in the balance sheet.

Academics and experts in finance argue that Cash flow statements prepared using the indirect method suffer from a serious drawback: Since this method reports only the net cash flow from operations, it does not allow a comparison and analysis of operating cash inflows and outflows by function with the revenues and expense activities that generated them. Hence it provides little new information about the specific components of a firm’s cash generating ability.

As a majority of firms provide only the indirect method statement, it is often necessary to convert an indirect one into a direct one. The main advantage of **the direct method** is that it shows operating cash collections and payments and is therefore more consistent with the basic purpose of the statement – to provide information about company's cash receipts and payments. Advocates of the direct method argue that information about the specific sources of operating cash receipts and the purposes for which operating cash payments were made is more useful to users than the net cash flow from operating activities.

Under the direct method, net cash from operations is computed by adjusting each item in the income statement from the accrual basis to the cash basis. An efficient way to apply the direct method is to adjust each revenue and expense reported in the income statement in the order in which it is listed. Table 4 shows a direct method statement of cash flows for J&J Company. We use the data from Table 1 (Comparative Balance Sheets) and Table 2 (Income statement) to calculate the following:

Cash Collections: The income statement reported revenues from customers of \$61,587,000. But in order to determine how much of it was collected in cash, we have to look at the change in accounts receivable during the year. Since it increased by \$207,000, this much revenue did not translate into cash. Thus, it will be deducted from sales revenue to arrive at the amount of cash collected from customers. Net income also reports interest collection of \$107,000. Finally there is “Other net income” of \$768,000, which could be gain from sale or purchase of business or other assets.

Cash payments: J&J income statement reported cost of goods sold of \$18,792,000. To determine cash payments to suppliers, we deduct increase in accounts payable of \$20,000, since the company paid less cash than the expenses incurred. Similarly we deduct a portion of the total depreciation expense from cash payment to suppliers, since it is a non-cash charge. Since the indirect method cash flow statement reported a total depreciation and amortization expense of \$2,939,000, it has to be added back to the expenses to arrive at the total cash payments made during the period. We prorated the depreciation amount between suppliers and employees, in the absence of any further information. However, it does not influence the net cash provided by operations. Finally, an increase in inventories of \$196,000 implied that purchases exceeded cost of goods sold. As a result, it will be added to cost of goods sold to arrive at cash purchases.

The company reported operating expenses of \$19,424,000 in the income statement. To determine the cash paid for operating expenses, we must adjust this amount for any changes in prepaid expenses and accrued expenses payable. A decrease in prepaid expenses of \$224,000 implies that cash paid for operating expenses was lower than operating expenses reported in income statement. Hence, it should be deducted from operating expenses. However, the indirect method statement combines this adjustment with a pool of others and reports an increase in “Other non cash items” of \$626,000. It is virtually impossible to determine what is hiding inside that number.

Table 4 Direct method: (J&J)

Item	Income statement	Adjustment	Cash
Cash collected from customers			
Net sales	61,587.0		
Increase in accounts receivables		−207.0	61,380.0
Cash received from other sources	768.0		768.0
Cash collected from interest	107.0		107.0
Total cash collected	62,462.0		62,255.0
Cash payments			
To suppliers			
COGS	18,792.0		
Increase in inventory		196.0	
Increase in accounts payable		−20.0	
Depreciation		−1,445.2	
Total payments to suppliers			−17,522.80
To employees			
Selling and administrative expenses	19,424.0		
Litigation gain		−50.0	
Non cash items		−626.0	
Depreciation		−1,493.8	
Total payments to employees			−17,254.2
For R&D	6,844.0		
Net cash paid for “other assets”		501.0	
Total cash payments for R&D			−7,345.0
For interest	528.0		
Interest payable		−37.0	
Total cash payments for interest			−491.0
For income taxes	3,613.0		
Deferred taxes		−356.0	
Total cash payments for taxes			−3,257.0
Total cash payments	49,201.0		−45,870.0
Net cash provided by operating activities	13,261.0		16,385.0

The R&D expense of \$6,844,000 is adjusted for \$501,000 to reflect a net increase in assets that were paid in cash but were not expensed. The SCF prepared by the indirect method shows an increase in “Other Assets” of \$574,000. The company’s footnote to income statement shows an interest capitalized of \$73,000. Since capitalized interest is a source of cash, we subtracted this amount before taking the net amount ($\$574,000 - \$73,000 = \$501,000$) as an increase in cash outlay on some long-term assets.

The income statement reported an interest expense of \$528,000, but the cash expense on interest was only \$491,000, as reported at the bottom of the indirect method statement. Hence, the difference must be an increase in accrued interest payable of \$37,000, which will be deducted from operating interest expense to arrive at the cash payments for interest.

Finally, the income-tax expense of \$3,613,0,000 needs adjustment to arrive at the amount of cash taxes paid. The indirect method statement reports that the deferred taxes increase by 356,000, meaning that the company did not pay this amount in cash. If we subtract this from income tax expense, we arrive at \$3,257,000 as cash payment for taxes. However, at the bottom of the indirect method statement provided by the company, we are told that the company paid only \$2,442,000 for taxes, implying that an additional \$815,000 has been added to cash by non-payment of some taxes. It is hard to deduce the company's intention behind this massive aggregation. Sometimes, a careful reading of footnote data may reveal additional information on aggregated accounts.

It should be noted that the net cash of 16,385,000 provided by operating activities under the direct method is the same as the number in the indirect method statement.

72.4 Problems with Cash Flow Statements

An operating cash flow Statement has long been considered a reliable barometer of a firm's financial health and less susceptible to creative accounting than its cousins the income statement and the balance sheet. However, recent corporate scandals as well as some research suggest that cash flow statements may also be susceptible to manipulation by corporate managers.¹⁰

The events surrounding the accounting debacles of Adelphia,¹¹ Dynegy, Qwest, Tyco,¹² WorldCom, and others demand a new look at this key financial statement, particularly because these cases reflect problems in reporting of cash flows by these corporations. Certain accounting shenanigans have been used to artificially boost reported cash flow or mask declines in operating cash flow.¹³

As further evidence of potential manipulation, the Securities and Exchange Commission recently forced seven well-known companies to amend their financial statements. Harley-Davidson, Caterpillar, and General Motors were among the companies affected by the ruling. Their 2002 and 2003 financial statements classified money owed to them by dealerships as an outstanding investment, meaning that the operating cash flows they reported were artificially high. Harley's 2003 cash flow was 29% lower after the restating; Caterpillar's tumbled by a staggering 372%. At GM, these reclassifications reduced its operating cash flow from \$7.6 billion to \$3 billion.¹⁴

¹⁰See for example, Sender (2002).

¹¹See Solomon, Deborah (2002).

¹²See Maremont (2002).

¹³"For example, after WorldCom's reverse-engineering subterfuge, many have learned to look for excessive capitalization of cash expenditures. Others now scrutinize the cash flow statements for non-recurring sources of cash, such as receipt of an income tax refund and securitization of payables." Mark A. Siegel, "Accounting shenanigans on the cash flow statement" March 2006.

¹⁴Judith Burns, "SEC tells Automakers to re-tool cash flow accounting", Wall street Journal Online (February 28, 2005).

Research out of Georgia Tech's business school has also found that a number of prominent companies routinely "miscategorize money due from vendors, such as members of their distribution network. Typically, firms record the IOUs as part of their cash flow from investments rather than as reductions to their operating cash flow."¹⁵ Other practitioners and researchers have also pointed out to the existing flaws in rules regarding cash flow statements.¹⁶

The accounting problems brought to light raise important questions about current accounting standards for the statement of cash flows and how they are being applied: Is the statement as useful to investors and creditors as it should be?¹⁷ Can it and should it be improved? Would improvements in the statement increase investors' and creditors' confidence in financial accounting and reporting? We will classify our criticism of existing rules under three categories: Problems due to the arbitrariness of classifications of cash flows of different types, problems due to the method chosen for the preparation of the SCF and problems due to a possible manipulation of several items reported in the SCF.

72.4.1 Arbitrariness of Classifications

Most classification systems are somewhat arbitrary. Classification involves weighing the relative significance of similarities and differences. Under SFAS-95, the unavoidable arbitrariness of the three-way classification sometimes results in classifying similar cash flows differently and different cash flows similarly. Additionally, although the three-way classification of cash flows under SFAS-95 links certain related cash flows, it does not link other cash flows that are either functionally or legally related.

72.4.1.1 Operating Versus Investing Activities

Under the current rules, FASB classifies the purchases of inventory as operating activities and purchases of plant assets as investing activities largely because "payments to suppliers for inventories are regarded normally as part of operating activities and payments for property, plant, and equipment as part of investing activities."

¹⁵"SEC acts to curb cash flow shenanigans" by Darren Dahl, Inc.com, June9, 2005.

¹⁶See for example, Cash flow myths. By: Rosen, Al, Canadian Business, 00083100, 3/12/2007, Vol. 80, Issue 6, "We tell our institutional money-managing clients that all cash flow mistakes result from the arbitrary timing or classification of management actions or inactions. The cash flow statement represents cash amounts. But, they are unevenly categorized, and worst of all, smoothed and shifted from period to period. . . . Investors tend to forget that accounting cash flows are not "real." Actual cash flows into a company are first recorded using an accrual (non-cash) income basis, and then translated back into a quasi-cash basis on the cash flow statement. This back-and-forth dance creates what I like to call leakage. Leakage can be caused by pension plans, income taxes, stock options, corporate acquisitions, minority interests, asset securitizations and a host of other factors."

¹⁷See for example, Livnat and Zarowin (1990).

During the past 20 years, many researchers have argued that the purchase and sale of inventories is fundamentally the same as the purchase and sale of plant assets; both are usually considered part of a firm's normal operating activities. The case for treating the purchase and sale of plant assets differently from the purchase and sale of inventories must be based on the grounds that they are of different significance to those interested in the cash inflows and cash outflows of a business enterprise. Although the purchase and sale of plant assets are of special significance because they are of relatively infrequent occurrence, because they are often relatively large in amount, and because management is likely to have more control over their timing than it does over the purchase and sale of inventories, Nonetheless, distinguishing between operating and investing activities is likely to be troublesome.

Thus, to the extent that payments for inventories and plant assets are fundamentally alike, the distinction between operating and investing activities is flawed or unnecessary. This notwithstanding, a consensus has developed that payments for inventories should be classified as operating outflows whereas payments for plant assets should be classified as investing outflows, consistent with the finance literature.

72.4.1.2 Operating Versus Financing Activities

Although the accounting profession talks about the importance of comparability of financial information, the fact that interest paid is treated as an operating activity while dividends paid is treated as a financing activity makes it difficult, if not impossible, to compare the performance of companies that make different financing choices.

Clearly, interest and dividend collections result from making loans or investments in debt and equity securities and, therefore, relate to investment activities. Similarly, interest payments result from incurring debt and, therefore, relate to financing activities. Indeed, the finance literature emphasizes the similarity of interest and dividend payments – interest is paid for the use of debt capital, whereas dividends are paid for the use of equity capital. It follows, therefore, that interest and dividend receipts are not operating inflows and interest payments are not operating outflows, notwithstanding SFAS-95.

The FASB acknowledges that a reasonable case could be made for classifying interest payments as financing outflows and interest and dividend collections as investing inflows. But the FASB notes that traditionally, interest collections and interest payments have been classified as operating flows. Additionally, it perceives widespread support for what might be referred to as the inclusion concept, whereby "... operating cash flow should, insofar as possible, include items whose effects are included in determining net income to facilitate an understanding of the reasons for the differences between net income and net cash flow from operating activities." Because of this inclusion concept, the FASB concludes that it "... was not convinced that changing the prevalent [prior] practice ... would necessarily result in a more meaningful presentation of cash flows."

It is inconsistent, however, for the FASB to cite the inclusion concept and prior practice to decide that cash flow from operating activities should include interest payments and interest and dividend collections but exclude cash flows relating to

gains and losses from plant asset disposals. Like interest payments and interest and dividend collections, gains and losses on plant asset disposals are traditionally included in net income, and the cash flows relating to these gains and losses have been included in operating funds flow under prior practice.

By classifying interest and dividend collections and interest payments as operating cash flows, the FASB has foregone the opportunity to more substantially improve the quality of financial reporting by sharpening the distinction among operating, investing, and financing cash flows. It should reconsider the issue.

A similar ambiguity arises in classifying payments of amounts borrowed. Cash outflows to suppliers and lenders are fundamentally alike, yet the former are classified as operating whereas the latter are classified as financing under SFAS-95. It is generally argued that some activities with a financing dimension, such as inventory purchases on account, should be excluded from financing activities because their spontaneous financing components are incidental to an enterprise's operations and do not affect its capital structure. In its 1980 Discussion Memorandum, the FASB also notes this ambiguity in the three-way classification. Nevertheless, FASB still favors classifying cash outflows to suppliers as operating and cash outflows to lenders as financing.

There are at least two inconsistencies in the reasoning underlying this classification. First, it assumes that inventory purchases on account always involve current liabilities. Second, it assumes that report users are interested principally in the capital structure, which excludes current liabilities. What about those admittedly less common situations where inventory purchases on account involve long-term liabilities? What about those report users interested in the financial structure, which includes all liabilities, not just long-term liabilities?

Because cash outflows to suppliers and lenders are fundamentally alike, the distinction between operating and financing activities based on payee is ambiguous. This ambiguity has been resolved by classifying outflows to suppliers as operating and outflows to lenders as financing, consistent with the finance literature.

Another ambiguity in classification involves "**capital leases**." It is true that, on day one, a lease is a noncash transaction. However, the payment for the lease principal and interest in later days is undoubtedly a cash flow event in a single transaction. And yet, the payment for the principal is treated as a financing activity, while the payment for the interest is shown as an operating activity. One transaction is broken into two parts; it has lost its wholeness. It would be more informative if both were brought together as one transaction under financing activities.

72.4.1.3 Investing Versus Financing Activities

SFAS-95 classifies cash flows according to the transactions that cause them. Cash inflows are not classified according to uses and, similarly, cash outflows are not classified according to sources. For example, cash receipts from the sale of plant assets are investing inflows even when the proceeds are used to retire debt, and cash receipts from issuing debt are financing inflows even when the proceeds are used to purchase plant assets. Such cash inflows and outflows might be viewed as functionally related, but are not necessarily "linked" and are not reported in the same section of the cash flow statement under SFAS-95.

72.4.2 Direct Versus Indirect Method Controversy

As mentioned before, companies favor the indirect method of preparing SCF, even though FASB recommends use of the direct method. The main advantage of the direct method is that it shows operating cash collections and payments and is therefore more consistent with the basic purpose of the statement – to provide information about company's cash receipts and payments.¹⁸ Indirect method does not report operating cash receipts and payments. Advocates of the direct method argue that information about the specific sources of operating cash receipts and the purposes for which operating cash payments were made is more useful in determining a company's future ability to generate cash flows to meet its obligations, to reinvest in its operations and to pay dividends to its owners.

Most companies argue that their accounting systems do not collect information in a manner that allows them to determine amounts such as cash received from customers or cash paid to suppliers. They prefer to use the indirect method since it is less costly to adjust net income to net cash flow than it is to report operating cash receipts and payments. Companies have lobbied against the direct method, urging FASB to adopt the indirect method.¹⁹

It is ironical that the statement of cash flows under either method does not meet the needs of the finance community, which is often interested in knowing how much cash was taken out of the business in a given year.²⁰

Finance textbooks and academicians argue that a cash flow statement should be able to summarize the cash activities of a firm much the same way as income statement summarizes the accrual activities of a firm.

Since the balance sheet identity tells us that the value of a firm's assets is equal to the value of its liabilities plus the value of its equity, the cash flow from a firm's assets must be equal to the cash flow to suppliers of capital to the firm. Thus the basic cash flow identity is:

$$\text{Cash flow (CF) from assets} = (\text{CF) to creditors} + (\text{CF) to stockholders}.$$

This identity reflects the fact that a firm generates cash through various activities and that cash is either used to pay creditors or paid out to the owners of the firm.

Cash flow from assets involves three components: operating cash flow, capital spending and change in working capital. Operating cash flow is defined in a slightly

¹⁸Orpurt and Zang (2009) show that direct method leads to a better forecast of future operating performance and cash flows.

¹⁹Accounting Trends and Techniques published by the American Institute of Certified Public Accountants (AICPA2005) reports that of 600 U.S. public firms surveyed over 2001–2004, only seven or eight firms disclosed DM statements of cash flows each year. Because few firms were expected to use the DM, FASB requires DM disclosing firms to also provide an IM schedule for comparability (SFAS No. 95, paras. 29–30). However, no separate DM schedule is required with IM statements. Miller and Bahnson (2002) argue that dual reporting required with DM disclosure may cause managers to prefer the IM.

²⁰See Ross et al. (2010).

different way under finance literature: Expenses associated with firm's financing of its assets, such as interest, are deducted because they are not operating expenses. Depreciation, being a non-cash expense, is added back to earnings.

Capital spending refers to the net spending on fixed assets (purchase of fixed assets minus sale of fixed assets).

Lastly, change in working capital is measured as the net change in current assets relative to current liabilities and represents the amount spent on working capital.

In order to calculate sources and uses of cash in a business, we need to trace the changes in firm's balance sheet to see how the firm obtained and spent its cash during same period. For example in the balance sheet for J&J, net fixed assets increased by 7,766 and its inventory, for all the three categories together, increased by \$198. (All figures are in millions.) Where did the cash come to pay for these purchases? To answer this and related questions, we need to identify those changes that brought in cash (sources) and those that used up cash. Logically, if an asset (liability) account increases (decreases) over the period, the firm must have used cash. Similarly, if an asset (liability) account went down (up) over the period, then the firm must have sold some assets (undertaken new liability) – a source of cash. Table 1, column 3, shows the change in each item on the balance sheet and the direction of the change. An increase in asset, for example, inventory, is a net use of cash because the firm paid to get additional inventory. An increase in Accounts receivable of \$135 is also a use of cash because the company has effectively given a loan receivable by the end of the year. By the same logic, increase in accounts payable of \$82 is a source of cash because the company borrowed this amount from its suppliers for a short term. Accrued liabilities, on the other hand, went down by \$525, so J&J must have paid off \$525 worth of short-term debts- A use of cash. Base on this discussion we can summarize the sources and uses of cash from the balance sheet as follows (Table 5):

Notice that the net addition to cash is same as the \$3,545 change on the balance sheet. This statement, although simple, tells us what happened during the year. The company increased its total liabilities by a lot but also acquired a lot of assets. However, it still does not tell the whole story. For example, it does not explain why the increase in accounts payable figure in the balance sheet does not match with the number reported in the indirect method (Many other numbers like increase in inventory, accounts receivable, deferred taxes and depreciation are also different in the two statements). Second, it would be more enlightening to have more details reported. For example, the increase in retained earnings is net income (a source of funds) less dividend (a use of funds) but we don't see the breakdown. Finally, we see only an increase in net fixed assets. Total or gross spending would be more interesting to know.

72.4.3 Manipulation and Distortion of Cash Flow Statements

Recent focus on SCF has revealed that Companies can either artificially boost reported cash flows or present unsustainable cash flows or mask declines in operating cash flows by using many of the loopholes in the existing rules.

Table 5 Changes in Cash Flow (J&J)

	2011	2010	Change
Sources of cash			
<i>Increase in liabilities</i>			
Loans and notes payable	7,617,000.00	6,318,000.00	1,299,000.00
Accounts payable	5,623,000.00	5,541,000.00	82,000.00
Accrued rebates	2,512,000.00	2,028,000.00	484,000.00
Income taxes	578,000.00	442,000.00	136,000.00
Deferred taxes	1,447,000.00	1,424,000.00	23,000.00
Long-term debt (Note 7)	9,156,000.00	8,223,000.00	933,000.00
Other liabilities	6,567,000.00	5,947,000.00	620,000.00
Retained earnings	77,773,000.00	70,306,000.00	7,467,000.00
<i>Decrease in assets</i>			
Raw materials	1,073,000.00	1,144,000.00	-71,000.00
Deferred taxes	2,224,000.00	2,793,000.00	-569,000.00
Deferred taxes	5,096,000.00	5,507,000.00	-411,000.00
Prepaid and others	2,273,000.00	2,497,000.00	-224,000.00
Doubtful account	-340,000.00	-333,000.00	-7,000.00
Depreciation	-15,873,000.00	-14,492,000.00	-1,381,000.00
Amortization of intangible	-5,509,000.00	-4,857,000.00	-652,000.00
Total sources			14,359,000
Uses of cash			
<i>Increase in assets</i>			
Marketable securities	8,303,000.00	3,615,000.00	4,688,000.00
Accounts receivables – gross	10,114,000.00	9,979,000.00	135,000.00
Goods in process	1,460,000.00	1,395,000.00	65,000.00
Finished goods	2,845,000.00	2,641,000.00	204,000.00
Land improvement	738,000.00	714,000.00	24,000.00
Buildings	9,079,000.00	8,863,000.00	216,000.00
Machinery and equipment	18,032,000.00	17,153,000.00	879,000.00
Construction	2,577,000.00	2,521,000.00	56,000.00
Goodwill	15,294,000.00	14,862,000.00	432,000.00
Intangible – gross	22,225,000.00	21,180,000.00	1,045,000.00
Other assets	3,942,000.00	3,690,000.00	252,000.00
<i>Decrease in liabilities</i>			
Accrued liabilities	4,100,000.00	4,625,000.00	-525,000.00
Accrued compensation and employee relate	2,642,000.00	2,777,000.00	-135,000.00
Employee Obligations	6,087,000.00	6,769,000.00	-682,000.00
Comprehensive Income	-3,531,000.00	-3,058,000.00	-473,000.00
Treasury stock	-20,783,000.00	-19,780,000.00	-1,003,000.00
Total uses			10,814,000
Net addition to cash			3,545,000.00

The following examples show how companies can employ certain techniques (many of which are within GAAP) to show improved reported cash flows.

72.4.3.1 Manipulation of Payables

The easiest thing that companies can do to improve reported operating cash flow is to slow down the rate of payments to their vendors. A direct consequence of this policy is to boost the reported growth in cash flows from operations. If analysts or investors expect the current period improvement to continue, they may be mistaken; vendors will eventually put increasing pressure on the company to pay more timely. Therefore, any benefit may be unsustainable or, at minimum, any year-over-year improvement in operating cash flow may be unsustainable.

The extension of payables can be identified by monitoring number of days' sales in payables (DSP).²¹ This metric is calculated as the end-of-period accounts-payable balance divided by the cost of goods sold and multiplied by the number of days in the period. As DSP grow, operating cash flows are boosted.

A look at J&J accounts payables reveals that the company has been experiencing huge fluctuations in this figure over the years: Its accounts payable increase by a modest \$20,000 in the current year, although it went down by \$507,000 and \$272,000 in the previous 2 years. The figures show that while the company received a significant benefit to cash flows from operations in 2009 and 2010, that benefit began to slow in subsequent periods, indicating that J&J will probably be unable to continue to fuel growth in operating cash flow using this method.

Another important point to note is that the firm shows an increase of \$20,000 in accounts payables in the SCF, the balance sheet shows an increase of \$82,000. It is surprising that one of the most popular adjustments to accrued expenses is not shown clearly and accurately.

72.4.3.2 Financing and Securitization of Payables

A more complicated version of stretching out payables is the financing or securitization of payables. This occurs when a company uses a third-party financial institution to pay the vendor in the current period, with the company then paying back the bank in a subsequent period. In recent years, increasing concern on keeping debt off the balance sheet has led companies to sales of receivables, with and without recourse. Economically, these activities still represent a form of financing. The accounting rules, however, allow these transactions to be treated as operating activities instead of financing activities. This makes these transactions doubly attractive to companies: They keep the borrowing off the balance sheet and inflate cash provided by operations. Theoretically, the pledge of accounts receivable is the same as the sale of accounts receivable, but they are treated differently in the cash flow statement. This treatment is inconsistent.

An interesting corollary to the impact on operating cash flow from securitizations is the impact on earnings. Specifically, in many cases companies can report gains

²¹ For a detailed analysis of problems with payables, see Nurnberg (2006).

when long-term accounts receivable are securitized. This occurs because the book value of the receivables at the time they are securitized does not include all the future interest income that is to be earned. Yet the entity purchasing the receivables will have to pay for that interest. As a result, because the amount received for the receivables is greater than the book value, a gain is generated. The amount of the gain can be affected as well by a variety of management assumptions, such as the expected default rate of the receivables securitized, the expected prepayment rate, and the discount rate used.

GAAP does not prescribe where on the income statement this gain is to be recorded. While one company may report the gain on sale of the receivables within revenues (the most aggressive approach), another might record it as an offset to selling, general, or administrative expenses. Another company might report the gain “below the line” in other non-operating income.

72.4.3.3 Deferred Employee Compensation

Another serious distortion involves deferred employee compensation. Many companies pay deferred compensation in the form of stock options that are off the balance sheet when issued. When the company later redeems the options by paying cash, the move is treated as a financing activity – as though it were a capital allocation, when in reality this payment is not any different from wages. As such, it should be treated as an operating activity.

72.4.4 Stock Buybacks

A second issue related to stock options that affects reported cash flows is the buyback of company stock. A large number of companies have, in recent periods, been buying back their own stock on the open market. In a majority of cases, this activity is due to stock-option activity. Specifically, when the price of a company’s stock rises, many of those who hold stock options exercise those relatively cheap options. If companies do nothing to offset the larger number of outstanding shares that exist as a result of the growing number of in-the-money options, earnings per share can be negatively affected. Management of such companies, therefore, face a choice: They can allow earnings per share to be diluted by the growing share count or they can buy back company stock to offset that dilution.

From an accounting standpoint, the impact of options on the income statement is usually minimal, as discussed above. On the cash flow statement, the tax benefit of option exercises is a source of operating cash flow, benefiting those companies whose option exercises grow. Cash expended by a company for the buyback of corporate stock, however, is considered a financing activity on the cash flow statement. Consequently, as option exercises grow, so does the boost to operating cash flows for the tax benefit, but the outflows for stock buybacks to offset dilution of earnings are recorded in the financing section of the cash flow statement.

Interestingly, as a company’s stock price rises, more options are generally exercised and the company must buy back more stock at the ever-higher market

prices. In some cases, the entire amount of cash flow generated by operations in recent periods could be expended to buy back company stock to offset the dilution from in-the-money options. (J&J used \$27,970,000 to buy back its own shares. This amount is reported as a financing outflow.) Therefore, when analyzing the true earnings power of a company as measured by cash flows, it is important to consider the cash expended to buy back stock to offset dilution. This cash outflow should be subtracted from the operating cash flow in order to calculate the true free cash flow the company generated in the period in question.

72.4.4.1 Marketable Securities

For many years, the distinction between cash equivalents (investments having no principal risk) and other marketable securities has caused serious confusion for the untrained reader of financial statements. Cash equivalents are treated as part of cash, while marketable securities are shown as investing activities. Where a company's portfolio manager does a lot of trading and switching between these two types of securities, very large numbers will appear in the investing activities section as "purchases of marketable securities" and "sales of marketable securities." It is not uncommon to see financial statements in which these numbers represent the largest cash flows. In J&J's balance sheet, marketable securities represent the largest cash outflow: an increase of \$4,688,000. Similarly, the SCF reports "Sale of investments" at \$11,101,000, which is classified as cash inflow under investing activity. If an untrained reader is asked what the most significant events for the company were during the year, the answer might be the purchasing and selling of marketable securities. Yet these numbers are irrelevant for understanding the company's performance. They merely clutter up the statement and cause confusion. Because SFAS 102 requires the buying and selling of securities to be treated as operating activities, the situation will be even worse.

72.4.4.2 Dividend from Affiliated Companies

Another problem that causes distortions in operating cash flows stems from the treatment of dividends received as an operating activity rather than as an investing activity. When a company has significant investments in affiliated companies, it has the ability to manipulate its own "cash provided by operations" by increasing the dividends it receives from such companies. Furthermore, dividend income comes from investments. If the former is shown in the section of operating activities, and the latter placed in investing activities, the reader of financial statement will not be able to visualize the whole picture of investment strategy.

72.4.4.3 Income Taxes

Another potentially serious distortion to operating cash flows comes from the rule that requires taxes to be treated as an operating activity, even when the gain being taxed is included in investing activities. For example, consider a company that has low operating profit but has a large gain from the sale of investments. The bulk of the pretax income is from this gain, and therefore the bulk of the income tax expense is related to this gain. On the statement of cash flows, however, the gain is removed

from operating activities and included under investing activities instead, as part of the proceeds from the sale of the investments. But the income tax expense on that gain remains in the operating activities sections, generating substantial negative cash from operations. This clearly is misleading, and violates the matching rules required on the income statement.

Because income tax payments and refunds are operating cash flows under SFAS-95, cash flow from operating activities includes the income tax effects of gains and losses relating to investing or financing activities, such as gains and losses on plant asset disposals and similar items.

Income taxes should be allocated in the cash flow statement in order to more accurately distinguish among cash flows from operating, investing, and financing activities.

Under the current rules of SFAS-95, when the amounts are material, it is recommended that firms disclose the income tax effects of gains and losses relating to investing and financing activities, to alert report users that the reported cash flow from operating activities is contaminated by the tax effects of these gains and losses. The J&J cash flow statement reports income taxes under current assets, current liabilities as well as in long-term assets. Although no detail is provided, it is possible that the figures pertain to income tax effects of gains and losses in purchase of businesses or sale of assets.

72.4.4.4 Sale and Purchase of Assets

A major problem with the investing activities section is that it is based on the rule that only cash amounts may be shown in investing and financing activities. Thus, for example, if one company acquired another at a cost of \$10 billion, but only \$1 billion of it was in cash, with the rest paid in the form of debt and equity instruments, the cash flow statement would show only the \$1 billion cash amount paid as the **cost of the acquisition**. The other \$9 billion would be relegated to a footnote. The untrained reader would get a false picture of the true cost of the acquisition. This is another example showing the deficiency of the current rules in preparing the cash flow statement. The rules ignore the vision of a complete transaction.

Moreover, there are certain inconsistencies in reporting installment purchases and sales of plant assets under SFAS-95. For installment purchases of plant assets, only early payments of principal, such as advance payments and down payments, are investing cash outflows; subsequent installment payments of principal are financing outflows. For installment sales of plant assets, however, all installment receipts of principal are investing inflows – not just the early receipts of principal.

The inconsistency results from classifying all installment receipts of principal on sales of plant assets as investing inflows but only some installment payments of principal on purchases of plant assets as investing outflows. Moreover, because most installment payments of principal on purchases of plant assets are classified as financing outflows, a comparative cash flow statement for a series of years does not report all the investing and financing activities of a firm.

72.4.4.5 Non-cash Transactions

In the Exposure Draft to SFAS-95, the FASB proposed that the primary purpose of the cash flow statement is to provide information about the cash receipts and cash payments of a firm during a period, and that a secondary purpose is to provide information about the investing and financing activities of the firm during the period. Consistent with these two purposes, the Exposure Draft required disclosure of noncash investing and financing transactions either in the cash flow statement or in a separate schedule. Examples of noncash investing and financing transactions include: converting debt to equity, acquiring assets by assuming directly related liabilities, such as purchasing a building by incurring a mortgage to the seller, obtaining an asset by entering into a capital lease, and exchanging noncash assets or liabilities for other noncash assets or liabilities.

Unlike the Exposure Draft, however, SFAS-95 retains the primary purpose but deletes the secondary purpose of the cash flow statement. The FASB concluded that the effectiveness of the cash flow statement would be enhanced if its purpose was restricted to reporting only cash flows, and therefore prohibits reporting noncash transactions in the cash flow statement. Nevertheless, SFAS-95 requires disclosure of all noncash investing and financing transactions that affect recognized assets or liabilities but that do not result in cash inflows or outflows. The FASB reasons (that although these noncash transactions result in no cash inflows or outflows in the periods in which they occur, they generally have a significant effect on cash flows in subsequent periods. SFAS-95 makes more uniform the reporting of noncash investing and financing transactions by prescribing supplemental disclosure outside the cash flow statement and prohibiting their disclosure in the cash flow statement. Unfortunately, the result of this narrower focus is that a comparative cash flow statement, unlike a comparative statement of changes in financial position, does not report all of a firm's investing and financing transactions.

For example, if a firm acquires equipment under capitalized leases, the inception of the lease is a noncash investing and financing transaction to be disclosed outside the cash flow statement, whereas the subsequent lease payments are allocated between interest and principal and reported in the cash flow statement as operating and financing outflows, respectively. Accordingly, a comparative cash flow statement for a series of years does not report all investing and financing transactions relating to this lease. Neither the original investment outflow nor the original borrowing inflow is reported in the cash flow statement, but the repayment is reported later as a financing outflow. From a cash flow statement perspective, the firm appears to be paying off a phantom loan, although the loan itself is reported in the balance sheet.

A similar result occurs when a firm acquires real estate by assuming an existing mortgage or by giving a mortgage or installment note to the seller. Only the down payment is reported as an investing outflow; all subsequent principal payments are reported as financing outflows.

Still another example of the incomplete reporting of financing activities in the cash flow statement under SFAS-95 is the conversion or exchange of debt for equity. A comparative cash flow statement reports the original borrowing as a financing inflow, but settlement of the debt by conversion or exchange is not reported as a financing outflow because it does not involve cash.

72.4.4.6 Free Cash Flows

Another important deficiency in the current cash flow statement is the absence of the concept of “free cash flow.”

Free cash flow is defined as cash without any restrictions on its use. It is available for any purpose at any time. It is similar to the concept of un-appropriated retained earnings. Since earnings are usually appropriated for dividend payout, plant expansion, contingencies, and other needs, cash available after these appropriations would be more useful for management planning.

Free cash flow has become increasingly important in financial statement analysis, yet the accounting profession has ignored it. The authors suggest that the profession take a serious look at free cash flow, with a view to narrowing and standardizing the conceptual definitions. If the traditional cash flow statement can be extended to include a consistent concept of free cash flow, it would become far more informative and useful.

72.5 Conclusion

This chapter compares the distinctions in SFAS-95 among operating, investing, and financing activities with similar distinctions in the finance literature. It identifies several inconsistencies and ambiguities in the three-way classification of cash flows under SFAS95, and demonstrates how the effectiveness of the cash flow statement to distinguish among operating, financing, and investing activities is impaired.

More specifically, this chapter demonstrates that the three-way classification result in reporting similar cash flows differently and different cash flows similarly, just the opposite of the desired characteristic of financial reporting to report similar things similarly and different things differently. The paper pinpoints that the classification of interest, income taxes, and dividend collections as operating cash flows results in contaminating cash flow from operating activities with the interest and dividend effects of investing and financing activities as well as with the tax effects of gains and losses relating to investing and financing activities. Finally, this chapter highlights that the requirements to report third-party financing transactions and disclose noncash investing and financing transactions are ambiguous.

By highlighting these inconsistencies and ambiguities, it is hoped that readers will be more aware of certain limitations of cash flow statements prepared under SFAS-95. More specifically, it is hoped that readers have been made aware that the amount reported as cash flow from operating activities does not reflect all of the cash flow consequences of operating activities and, similarly, that the amounts reported as cash flow from investing and financing activities do not reflect all of the cash flow consequences of investing and financing activities, respectively. Accordingly, these amounts should not be incorporated into operating, investing and financing decision models without adjustment. We hope that FASB will soon recommend some important changes to make cash flow statements more transparent and useful to the investing community.

References

- Ali, A. 1994. The incremental information content of earnings, working capital from operations and cash flows. *Journal of Accounting Research* 32 (1): 61–74.
- American Institute of Certified Public Accountants (AICPA). 2005. Accounting trends and techniques. 59 Jersey City, NJ: AICPA21(1), 17–24.
- Barth, M.E., W.H. Beaver, J.R.M. Hand, and W.R. Landsman. 1999. Accruals, cash flows and equity values. *Review of Accounting Studies* 3: 205–229.
- Barth, M.E., D.P. Cram, and K.K. Nelson. 2001. Accruals and the prediction of future cash flows. *The Accounting Review* 76 (1): 27–58.
- Bowen, R., M.D. Burgstahler, and L.A. Daley. 1986. Evidence on the relationships between earnings and various measures of cash flow. *The Accounting Review* 61 (4): 713–725.
- Dechow, P.M. 1994. Accounting earnings and cash flows as measures of firm performance: The role of accounting accruals. *Journal of Accounting and Economics* 18: 3–43.
- Dechow, P.M., S.P. Kothari, and R.L. Watts. 1998. The relation between earnings and cash flows. *Journal of Accounting and Economics* 25: 133–168.
- Financial Accounting Standards Board. 1987. Statement of financial accounting concepts no. 1. Objectives of financial reporting by business enterprises. Stamford 1978.
- . 1995. Statement of financial accounting standards no. 95. Statement of cash flows. Nov. 1995.
- Finger, C.A. 1994. The ability of earnings to predict future earnings and cash flow. *Journal of Accounting Research* 32–2: 210–223.
- International Accounting Standards Board. 2004. Presentation of financial statements, revised in 2004.
- Livnat, J., and P. Zarowin. 1990. The incremental information content of cash-flow components. *Journal of Accounting and Economics* 13: 25–46.
- Maremont, Mark. 2002. How is tyco accounting for its cash flow? *Wall Street Journal* (5 March), C1-2.
- Miller, Paul B.W., and Paul R. Bahnson. 2002. Fast track to direct cash flow reporting. *Strategic Finance* 83–8: 51–57.
- Nurnberg, H. 2006. Perspectives on the cash flow statement under FASB Statement No. 95. Occasional paper series. *Center for Excellence in Accounting and Security Analysis*. New York: Columbia University.
- Ohlson, J., and J. Aier. 2009. On the analysis of firms' cash flows. *Contemporary Accounting Research* 26–4: 1091–1114.
- Orpurt, S., and R. Zang. 2009. Do direct cash flow disclosures help predict future operating cash flows and earnings? *The Accounting Review* 84–3: 893–935.
- Ross, S., R. Westerfield, and B. Jordan. 2010. *Fundamentals of corporate finance*. 9th ed. McGraw-Hill Irwin.
- Sender, Henry. 2002. Cash flow? It isn't always what it seems. *Wall Street Journal* (8 May): C3.
- Siegel, Mark. 2006. Accounting Shenanigans on the cash flow statement. *CPA Journal*, March 2006, 38–43.
- Solomon, Deborah. 2002. Adelphia overstated cash flow, revenue over past two years. *Wall Street Journal* (11 June): B5.

Do CEO Gender and Marital Status Affect Firm's R&D and Value? An Empirical Analysis Using Nonlinear Models

73

Keshab Shrestha, Cheng-Few Lee, and Abdul Ghafoor

Abstract

This chapter reviews some nonlinear models used in corporate finance research. We also propose and empirically test some hypotheses regarding the effects of CEO gender and marital status on firm's R&D investment and firm values. We find that there is no association between the R&D investment and CEO gender. Similarly, there is not an association between the R&D investment and CEO marital status. However, using the usual Tobin's Q to represent the firm value, we find that both single CEO and male CEO enhance the firm value. We also find a nonlinear relationship between firm value and each of the two variables like firm size and institutional ownership. Finally, we find a significant positive association between the free cash flows and firm value but a significantly negative association between the firm value and leverage.

Keywords

Nonlinear models · CEO gender · CEO marital status · Nonlinearity with interaction effect · Tobin's Q

K. Shrestha (✉) · A. Ghafoor

Department of Finance, School of Business, Monash University Malaysia, Bandar Sunway, Selangor, Malaysia

e-mail: keshab.shrestha@monash.edu; abdul.ghafoor1@monash.edu

C.-F. Lee

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: cflee@business.rutgers.edu

73.1 Introduction

In many corporate finance research, linear relationships are used to express association among different variables of interest. Furthermore, linear relationships are also used to estimate and test various theories in empirical studies. However, the relationships among variables are the outcome of optimization process and equilibrium conditions, which could in general lead to nonlinear relationships instead of linear relationships. Sometimes, a linear relationship is assumed or estimated to simplify the relationship. On the estimation side, such simplification has been done to save costs due to high computational costs in the past. Due to low computational costs as well as the development of symbolic computation, analyzing nonlinear models is more feasible now. Furthermore, linear relationships are special cases of nonlinear relationships, therefore, there is no cost to analyzing a possible nonlinear relationship.

Since Zerambka (1968), the function form papers in economics and finance research have started paying more attention to nonlinear models. For example, Lee (1976) has applied the functional form methodology to dividend policy to test for linear, log-linear, and nonlinear relationships. There are many other studies that use nonlinear models in corporate finance, and these studies are briefly discussed in the literature review section of this chapter.

In this chapter, we present four hypotheses regarding the effects of CEO's gender and marital status on a firm's R&D investment and firm values. We test these hypotheses using nonlinear models using the US data for 2008. Our empirical analysis indicates that there is no association between the R&D investment and CEO gender. Similarly, there is no association between the R&D investment and CEO marital status. However, there is significant association between the R&D investment and variables like firm value, free cash flows, and firm leverage. As to the hypothesis regarding firm value and CEO gender and marital status, we find some interesting results. Using the usual Tobin's Q to represent the firm value, we find that both single CEO and male CEO enhance the firm value. There are nonlinear relationships between firm value and each of the two control variables, firm size and institutional ownership. Finally, we find a significant positive association between the free cash flows and firm value, but a significantly negative association between the firm value and leverage.

73.2 Literature Review

This section presents a brief literature review on nonlinear models in corporate finance. Detailed information on some of the studies presented here can be found in Chen et al. (2004). Many studies use nonlinear models to examine the relationship between firm value (or firm performance) and ownership structure. Some examples of these studies include Jensen and Meckling (1976), Leland and Pyle (1977), Demsetz (1983), Fama and Jensen (1983), Stulz (1988), Morck et al. (1988), McConnell and Servaes (1990), Bagnani et al. (1994), Cho (1998), Holderness et al. (1999), Short and Keasey (1999), Himmelberg et al. (1999), Anderson and

Reeb (2003), Joh (2003), Florackis et al. (2009), Yu et al. (2012), and Hamadi and Heinen (2015).

Other studies analyze the nonlinear dependence of the effects of equity issuance on ownership structure (e.g., Wruck (1989), Hertz and Smith (1993), Fields and Mais (1994), Brous and Kini (1994), Mikkelsen et al. (1997), Slovin et al. (2000), and Aggarwal et al. (2002)). Cook et al. (2006) show that the greater the promotion efforts by investment bankers, prior to IPOs, the larger the underpricing of the issue. Similarly, Kim et al. (2010) examine whether the relationship between issue size and underwriting spreads is nonlinear and U-shaped.

Some studies use nonlinear models to analyze the corporate payout policy. These studies include Schooley and Barney (1994), Cook and Easterwood (1994), Alangar et al. (1999), Lie (2000), Setia-Atmaja et al. (2009). Lee et al. (2011) present a theoretical model that shows a nonlinear payout policy dependent on growth rate. Ham et al. (2020) analyze the information content of dividends regarding future earnings. They find a nonlinear effect of a dividend increase and decrease, whereby dividend increases have a more persistent effect on earnings compared to dividend decreases.

Nonlinear relations are also common in corporate capital structure research (e.g., Diamond (1991), Bagnani et al. (1994), Bae et al. (1994), Barclay and Smith Jr. (1995), Stohs and Mauer (1996), Roth and McDonald (1999), and Anderson et al. (2003)). Recently, Chen et al. (2021) used the interaction effect with a recession dummy variable and found that (i) debt maturity is pro-cyclical, (ii) higher-beta firms tend to have a longer maturity, and (iii) shorter maturity amplifies the sensitivity of credit spreads to aggregate shocks. Nonlinear models are also used in analyzing cash holdings (Opler et al. 1999) and corporate investment decisions (Allen and Cebenyoyan 1991; Thomas 1995; Martin 1996; Subrahmanyam et al. 1997; Denis et al. 1977a; Cho 1998; Ghosh and Ruland 1998). Recently, Bustamante (2016) has used the quadratic equation model to analyze corporate investments. Some studies use nonlinear models in antitakeover provisions research (e.g., McWilliams (1990) and Boyle, Carter and Stover (1998)).

Nonlinear models are also used to analyze the impact of board structure (Byrd and Hickman 1992; Denis et al. 1977b; Rosenstein and Wyatt 1997; Yermack 1996; Andres and Vallelado 2008; Coles et al. 2008). For example, Byrd and Hickman (1992) find a nonlinear relationship between bidding firms' abnormal returns and the proportion of independent outside directors on the board. Similarly, Denis et al. (1977b) found the probability of top executive turnover is significantly less sensitive to performance when officers and directors own between 5% and 25% of the firm's shares than when they own less than 5%.

Some studies use nonlinear models to analyze the relationship among price per share, dividend per share, and retained earnings per share for US industrial firms. Lee (1976) and Gilmore and Lee (1986) developed a generalized functional form (GFF) to analyze the relationship among the stock price, dividends, and retained earnings. They find that the linear form (sometimes even the log-linear form) is not a correct functional form for investigating the dividend effect of the electric utility industry. Moreover, the dividend effect is not significantly different from the retained earning

effect for 60% of the cross-section results after the true functional relationship has been determined. The GFF explicitly takes the interrelation between dividends and retained earnings into account; the functional form parameter also reflects the different relationships among the stock price, dividends, and retained earnings under various economic conditions. In sum, it has been established that the previous studies of determining the dividend effect of the electric utility industry may have pre-judged an important issue – the correct functional form. While linear and logarithmic forms are easy to handle, this consideration alone is not enough to justify the employment of either of these forms. In fact, it has been shown that there exists a generalized functional form, which allows a compact analysis of the effects of choice of functional form on determining the dividend effect of the electric utility industry.

Chang and Lee (1977) used two alternative techniques to analyze pooled time-series and cross-section data to test the importance of the firm effect and time effect in the financial analysis. Their techniques are also integrated with the functional form parameter estimation method to show the importance of appropriate functional form in handling a pooled time-series and cross-section type econometric model. The data on the electric industry show that both the time effect and cross-section effect are important in explaining stock price variation. It is also found that linear form (and/or) log-linear form is not always appropriate in testing the importance of both the time and firm effects in financial analyses.

The remaining part of this section is devoted to a discussion of the model specifications. Since there are numerous models, we will only discuss the model specifications not covered in Chen et al. (2004). When trying to understand whether a relationship among variables is linear or nonlinear, it would be helpful to use a specification where a possible linear relationship is a particular case of a nonlinear model. One of the ways to achieve this is to use the functional form as suggested by Zarembka (1968) and Lee (1976) using Box and Cox (1964) transformation $X^{(\lambda)}$ of variable X as defined below:

$$X^{(\lambda)} = \frac{X^\lambda - 1}{\lambda} \quad (1)$$

This transformation has the following property:

$$\lim_{\lambda \rightarrow 0} \frac{X^\lambda - 1}{\lambda} = \ln(X) \quad (2)$$

For example, Lee (1976) considers the following possible nonlinear model, depending on the value of the parameter λ ,

$$P_{it}^{(\lambda)} = \alpha_0 + \alpha_1 D_{it}^{(\lambda)} + \alpha_2 R_{it}^{(\lambda)} + \varepsilon_{it} \quad \varepsilon_{it} \sim N(0, \sigma^2) \quad (3)$$

where P_{it} , D_{it} , and R_{it} are the price, dividends, and retained earnings per share for the firm i and period t , respectively. The parameters α_0 , α_1 , α_2 , σ , and λ can be estimated

using the maximum likelihood estimator. It is important to note that the Eq. (3) becomes linear if λ equals 1 and log-linear if λ equals 0.

Alternately, one can use logistic transformation instead of Box and Cox (1964) transformation, where the logistic transformation of X is given by

$$\ln\left(\frac{X}{1-X}\right) \quad (4)$$

Lee et al. (2011) consider the following nonlinear model

$$\ln\left(\frac{X_{it}}{1-X_{it}}\right) = \alpha + \beta_1 Risk_{it} + \beta_2 D_{it}(g_{it} < cROA_{it})Risk_{it} + \beta_3 Growth_{it} + \beta_4 Risk_{it} \times Growth_{it} + \beta_5 \ln(Size)_{it} + \beta_6 ROA_{it} + e_{it} \quad (5)$$

Another specification is to use the smooth transition function instead of a piecewise regression. Christopoulos and Leon-Ledesma (2007) suggest using either the logistic smooth transition (LSTR) or exponential smooth transition model (ESTR). An example is given below for two independent variables (X and Z) case:

$$Y_t = \alpha + \beta_1 X + \beta_2 Z_t + \beta_3 F(Z_t : \gamma, c) + e_t \quad (6)$$

where for LSTR, the transition function $F(Z_t : \gamma, c)$ is given by

$$F(Z_t : \gamma, c) = \frac{1}{1 + e^{-\gamma(Z_t - c)}} \quad (7)$$

And, for ESTR, the transition function $F(Z_t : \gamma, c)$ is given by

$$F(Z_t : \gamma, c) = 1 - e^{-\gamma(Z_t - c)^2} \quad (8)$$

where c gives the location parameter and γ is the speed parameter.

73.3 Hypotheses

The main objectives of this chapter are to analyze the impact of characteristics like (i) CEO gender and (ii) CEO marital status on a firm's R&D investment and firm value measured by Tobin's Q. It is well recognized that managers are not a homogeneous group of individuals. According to the Upper Echelons Theory (Hambrick and Manson 1984; Hambrick 2007), managers are heterogeneous, and that firm-related decisions are affected by managers' experiences, values, and personalities.

There exists a body of research examining whether corporate outcomes can be influenced by more women on boards (e.g., Adams and Ferreira (2009), Ahern and Dittmar (2012), and Matsa and Miller (2013)). Some researchers argue that women tend to have a higher expectation of their responsibilities and that women have to demonstrate extra competence to reach executive positions (Fondas and Sassolos

2000; Eagly and Carli 2003; Dunn 2010). Researchers in Psychology have found evidence that indicates that women have better communication skills, and women tend to perform better in group problem-solving and decision-making tasks (Dallas 2002; Fondas 1997; Maznevski 1994; Robinson and Dechant 1997; Wood et al. 1985). We refer to this hypothesis that female CEOs are more responsible, more competent, better problem solvers, and better decision makers than male CEOs as the *competency hypothesis*. Furthermore, there exists literature on gender differences that suggest women are more conservative and risk-averse than men (see, e.g., Byrnes et al. (1999), Jianakoplos and Bernasek (1998), Johnson and Powell (1994), Levin et al. (1988), and Sunden and Surette (1998)). We refer to this argument as the *risk-aversion hypothesis*.

As to the marital status of the CEO, there is some interesting evidence. For example, Roussanov and Savor (2014) found that firms run by single CEOs exhibit higher stock return variance, use more aggressive investment policies, and do not respond to changes in idiosyncratic risk. Their findings hold when they use variation in divorce laws across states as an instrument for the CEO's marital status. This evidence indicates that the marital status itself drives choices rather than just reflecting innate heterogeneity in preference.

Now, we come to the hypothesis regarding R&D investment. Managers may prefer a quiet life (Hart 1983; Bertrand and Mullainathan 2003; Holmstrom 1979) and dislike the costly efforts associated with innovation projects like R&Ds (Chen et al. 2018). These arguments imply low investments in R&D, in general. However, when we compare a female CEO to a male CEO, the *competency hypothesis* implies that female CEO would invest more in R&D because they feel more responsible and are competent. However, the *risk aversion hypothesis* suggests that a female CEO would invest less in R&D, because R&D activities involve higher risk. Therefore, whether a female CEO would invest more or less in R&D compared to a male CEO is an empirical question. Therefore, we state our first hypothesis regarding CEO gender and R&D investment in the null form as follows:

Hypothesis 1: *There is no association between the CEO gender and R&D investment.*

As to the marital status of the CEO, Roussanov and Savor (2014) found that a single CEO invests more in R&D plus advertising compared to a married CEO. However, in the presence of CEO gender, we do not know the outcome. Therefore, we present the following hypothesis in the null form:

Hypothesis 2: *There is no association between the CEO gender and R&D investment.*

Regarding the firm value, the *competency hypothesis* implies that the firm value with a female CEO will be higher than the firm with a male CEO. On the contrary, the *risk-aversion hypothesis* implies that the female CEO may not take enough risk. Therefore, it will lead to a lower firm value. Therefore, the nature of the association

between the CEO gender and the firm value is an empirical question. We present the hypothesis regarding the CEO gender and firm value in the null form:

Hypothesis 3: *There is no association between the CEO gender and firm value.*

As to the CEO's marital status and firm value, one can argue that a single CEO will be able to take more risks because he or she does not have any family responsibilities. However, there is still the question that the risk a single CEO takes may be higher than the optimal level of risk for the firm. Therefore, we do not have a predetermined nature of relationship between the marital status of the CEO and firm value. Therefore, we present the last hypothesis in the null form.

Hypothesis 4: *There is no association between the firm value and the marital status of the CEO.*

In this chapter, we will test all four hypotheses using nonlinear models that will be presented in the next section.

73.4 Empirical Analysis

In this section, we estimate two nonlinear relationships between dependent variables and independent variables of interest. The first dependent variable is the R&D investment and the second dependent variable is the firm value. Recently, corporate finance researchers are interested in the impact of CEO characteristics like gender (Bugeja et al. 2012) and marital status (Roussanov and Savor 2014). Therefore, we will also empirically test the impact of CEO gender and marital status on the R&D investment and firm value. We specifically test the four hypotheses presented earlier.

In the cross-sectional analysis, we used the data for the fiscal year 2008 obtained from Roussanov and Savor (2014). The dataset consists of 908 firms. There are 213 firms with single CEOs, and the remaining 695 firms have married CEOs. Furthermore, there are 28 female CEOs and 880 male CEOs in our sample. All the continuous variables are winsorized at the top and bottom 1% level.

73.4.1 R&D Investments

In this section, we estimate the nonlinear dependence of R&D investment on the firm size. We will estimate three versions of the R&D investment function: (i) with no switching point, (ii) with one switching point, and (iii) with two switching points. In Eq. (9), only the linear effect of firm size is included. In Eq. (10), a piecewise linear function with one switching point (S_1) is used to express the nonlinear effect of firm size. Equation (11) shows a piecewise linear function with two possible switching points (S_1 and S_2).

With no switching points:

$$RDS = \beta_0 + \beta_1 EVS + \beta_2 Single + \beta_3 Male + \beta_4 INST + \beta_5 CFS + \beta_6 LTD + \beta_7 ADVS + \sum_{j=1}^{44} \beta_{7+j} IND_j + \theta SIZE + Error \quad (9)$$

With one switching point:

$$RDS = \beta_0 + \beta_1 EVS + \beta_2 Single + \beta_3 Male + \beta_4 INST + \beta_5 CFS + \beta_6 LTD + \beta_7 ADVS + \sum_{j=1}^{44} \beta_{7+j} IND_j + \theta_1 SIZE_1 + \theta_2 SIZE_2 + Error$$

$$RDS = \beta_0 + \beta_1 EVS + \beta_2 Single + \beta_3 Male + \beta_4 INST + \beta_5 CFS + \beta_6 LTD + \beta_7 ADVS + \sum_{j=1}^{44} \beta_{7+j} IND_j + \theta_1 SIZE_1 + \theta_2 SIZE_2 + Error \quad (10)$$

where

$$SIZE_1 = \begin{cases} SIZE, & \text{if } SIZE < S_1 \\ S_1, & \text{if } SIZE \geq S_1 \end{cases},$$

and

$$SIZE_2 = \begin{cases} 0, & \text{if } SIZE < S_1 \\ SIZE - S_1, & \text{if } SIZE \geq S_1 \end{cases}.$$

With two switching points:

$$RDS = \beta_0 + \beta_1 EVS + \beta_2 Single + \beta_3 Male + \beta_4 INST + \beta_5 CFS + \beta_6 LTD + \beta_7 ADVS + \sum_{j=1}^{44} \beta_{7+j} IND_j + \theta_1 SIZE_1 + \theta_2 SIZE_2 + \theta_3 SIZE_3 + Error \quad (11)$$

where

$$SIZE_1 = \begin{cases} SIZE, & \text{if } SIZE < S_1 \\ S_1, & \text{if } SIZE \geq S_1 \end{cases}$$

$$SIZE_2 = \begin{cases} 0, & \text{if } SIZE < S_1 \\ INSD - S_1, & \text{if } S_1 \leq SIZE < S_2 \\ S_2 - S_1, & \text{if } SIZE \geq S_2 \end{cases}$$

$$SIZE_3 = \begin{cases} 0, & \text{if } SIZE < s \\ SIZE - s, & \text{if } SIZE \geq s \end{cases}$$

The variable definitions are as follows:

<i>RDS</i>	=	R&D expenditures divided by sales
<i>EVS</i>	=	(market value of equity + book value of long-term debt – book value of total assets)/sales
<i>CFS</i>	=	Free cash flow divided by sales, where free cash flow equals operating income before depreciation – Interest expense – Taxes – Dividends
<i>LTD</i>	=	Long-term debt/(long-term debt + market value of equity)
<i>SIZE</i>	=	Log(market value of equity)
<i>ADVS</i>	=	Advertising expenses divided by sales
<i>INST</i>	=	Institutional holdings
<i>Male</i>	=	Male dummy = 1 if CEO is male
<i>IND_{1 to}</i>	=	Industry dummies
<i>IND₄₄</i>	=	
<i>Single</i>	=	Single dummy = 1 if CEO is single

The switching points are not assumed to be known in advance. They are rather empirically estimated. The locations of switching points are estimated using a grid search technique. The number and locations of switching points are chosen based on the log-likelihood test.

The summary statistics are presented in Table 1. The sample mean of EVS is positive with a negative median. This indicates that the distribution is far away from being symmetric. The mean for the variable Single is 0.235 as there are 213 firms with single CEOs out of 908 firms. Similarly, the mean for variable Male is 0.969 with 28 out of 908 firms having female CEOs. The average for RDS, CFS, LTD, ADVS, SIZE and INST are 0.034, 0.083, 0.269, 0.010, 7.233, and 0.787, respectively.

The correlation coefficients are presented in Table 2. The minimum correlation is –0.45 or negative 45%, which is between EVS and LTD. The maximum correlation is 38 percent which is between CFS and Size. The correlation between EVS and RDS is 25 percent. The correlation between RDS, one of the two dependent variables, and Single is 6%, and the correlation between RDS and Male is 1%. Similarly, the correlation between EVS and Single is 6%, and the correlation between EVS and

Table 1 Summary statistics: This table presents the mean, median, standard deviation, and the number of observations for each of the variables used in the chapter

	N	Mean	Median	Standard Deviation
EVS	908	0.083	−0.124	0.936
Single	908	0.235	0.000	0.424
Male	908	0.969	1.000	0.173
RDS	908	0.034	0.000	0.072
CFS	908	0.083	0.091	0.199
LTD	908	0.269	0.229	0.243
ADVS	908	0.010	0.000	0.023
SIZE	908	7.233	7.212	1.666
INST	908	0.787	0.828	0.172

Table 2 Correlation coefficient: This table presents correlation coefficients for the variables used in the study

	EVS	Single	Male	RDS	CFS	LTD	ADVS	SIZE
EVS	1.00							
Single	0.06	1.00						
Male	0.05	0.04	1.00					
RDS	0.25	0.06	−0.01	1.00				
CFS	0.24	−0.01	0.06	−0.15	1.00			
LTD	−0.45	−0.04	0.02	−0.26	−0.21	1.00		
ADVS	0.11	0.01	−0.05	0.01	−0.05	−0.03	1.00	
SIZE	0.24	−0.21	0.02	−0.03	0.38	−0.18	0.01	1.00
INST	0.19	0.08	0.01	−0.00	0.15	−0.12	−0.01	0.12

Male is 5%. Based on the two variables analysis, the CEO’s gender and marital status seem to affect the R&D and firm value. However, we will test the significance of these effects in the multivariate analysis to be presented later.

The results of the piecewise regression represented by three models (Eqs. 9, 10, and 11 are presented in Table 3. The log-likelihood test statistics are presented in Table 4. If we use the standard 5% level, then we conclude that there is one switching point for SIZE at 6.056. Therefore, rest of the conclusion is based on the model reported in Table 3 and Eq. (10), that is, the third column. The results indicate that there is no association between the R&D investment and the CEO’s gender. Similarly, there appears to be no association between the R&D investment and the marital status of the CEO. The coefficient of EVS is highly significant even at the 1% level. This is consistent with the high correlation between EVS and RDS. Since the coefficient of CFS is negative and significant at the 1% level, free cash flow seems to have negative association with RDS. Similarly, the coefficient of LTD is negative and significant at the 1% level implying that high levered firms spend less on R&D.

Table 3 Piecewise regression results: This table presents the result of the nonlinear piecewise regression represented by Eqs. (9), (10), and (11)

	No switching Eq. (9)	Single switching, Eq. (10)	Two switching, Eq. (11)
EVS	0.006*** (2.65)	0.007*** (2.75)	0.007*** (2.85)
Single	0.001 (0.18)	0.001 (0.17)	0.000 (0.07)
MALE1	-0.002 (-0.20)	-0.001 (-0.08)	-0.001 (-0.09)
INST	-0.015 (-1.36)	-0.006 (-0.55)	-0.011 (-0.94)
CFS	-0.062*** (-6.04)	-0.061*** (-6.05)	-0.061*** (-5.98)
LTD	-0.027*** (-3.03)	-0.028*** (-3.18)	-0.026*** (-2.98)
ADVS	0.069 (0.78)	0.067 (0.76)	0.084 (0.95)
SIZE	-0.001 (-0.58)		
SIZE ₁		-0.003*** (-2.64)	0.083*** (2.70)
SIZE ₂		0.002 (0.95)	-0.005** (-2.42)
SIZE ₃			0.003 (1.18)
Industry dummy	Yes	Yes	Yes
Constant	0.061* (1.79)	0.058* (1.75)	-0.209** (-1.99)
N	908	908	908
Adj. R-squared	0.4873	0.4908	0.4921

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 Log-likelihood ratio statistics: This table presents the log-likelihood value and log likelihood ratio test statistics used in determining the number of switching points

	Log-likelihood function	Likelihood ratio statistics
No switching point	1427.711	
Single switching point at 6.058 (SIZE)	1431.319	7.216***
Two switching points at 3.336 and 7.561 SIZE	1433.011	3.383*

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The log-likelihood ratio test statistic is given by $2[\ln(L_{i+1}) - \ln(L_i)]$ where $\ln(L_i)$ is the value of the log-likelihood function with number of switching points. The log-likelihood statistics have chi-squared distribution with degrees of freedom equal to 1

73.4.2 Firm Value

In order to test our third and fourth hypotheses, we estimate the nonlinear relationship between firm value (EVS) with two independent variables, SIZE and INST in the presence of other variables. Specifically, the following equation is estimated:

$$EVS = (\text{control variables}) + \beta_1 \text{Male} + \beta_2 \text{Single} + \theta_1 z_1 + \theta_2 z_1^2 + \theta_3 z_1^3 + \gamma_1 z_1 z_2 + \gamma_2 z_1 z_2^2 + \varphi_1 z_2 + \varphi_2 z_2^2 + \varphi_3 z_2^3 + e \quad (12)$$

where $Z_1 = \text{SIZE}$ and $Z_2 = \text{INST}$. The control variables used are *RDS*, *CFS*, *LTD*, *ADVS*, and the 44 industry dummies. We also analyze the impact of CEO gender and marital status. The results are presented in Table 5.

The results presented in Table 5 indicate that the coefficient of Single is positive and significant at the 10% level. Actually, the p-value is 0.052, which is almost significant at the 5% level. This implies that a single CEO has a positive impact on firm value. Similarly, the coefficient of Male is positive and significant at the 1% level. This implies a strong contribution of Male to the value of the firm. The nonlinear relationship between EVS and SIZE is only through the interaction effect with INST. But, INST clearly has a nonlinear relationship with EVS both through main effect and interaction effect.

In Table 5, we used EVS, which is firm value scaled by sales, as the dependent variable. However, the firm value is normally scaled by total assets and this definition is known as Tobin's Q, which is defined below:

Table 5 Nonlinear regression results with EVS: This table presents results of the estimation of Eq. (12) with EVS as the dependent variable. SIZE2 and SIZE3 represent SIZE raised to power 2 and 3, respectively. Similarly, INST2 and INST3 represent INST raised to power 2 and 3, respectively

Independent variables	Coefficient	T-Ratio	P-value
Single	0.119*	1.95	0.05
Male	0.415***	2.81	0.01
RDS	1.473***	3.02	0.00
CFS	0.754***	5.13	0.00
LTD	-1.211***	-10.07	0.00
ADVS	0.641	0.51	0.61
SIZE	0.135	0.29	0.77
SIZE2	0.053	0.82	0.41
SIZE3	-0.003	-0.86	0.39
INST	14.889***	3.50	0.00
INST2	-15.351**	-2.39	0.02
INST3	3.201	0.91	0.36
SIZE x INST	-1.516***	-2.87	0.00
SIZE x INST2	1.271***	3.18	0.00
Constant	-2.464*	-1.86	0.06
Industry dummy	Yes		
Adj. R-squared	0.388		
N	908		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6 Nonlinear Regression Results with Tobin's Q: This table presents results of the estimation of Eq. (12) with Tobin's Q as the dependent variable

Independent variables	Coefficient	T-ratio	P-value
Single	0.153***	3.12	0.00
Male	0.243**	2.05	0.04
RDS	0.448	1.15	0.25
CFS	0.356***	3.02	0.00
LTD	-1.087***	-11.27	0.00
ADVS	1.585	1.57	0.12
SIZE	-0.453	-1.22	0.22
SIZE2	0.121**	2.33	0.02
SIZE3	-0.006**	-2.47	0.01
INST	8.764**	2.57	0.01
INST2	-8.859*	-1.72	0.09
INST3	1.485	0.53	0.60
SIZE x INST	-0.892**	-2.10	0.04
SIZE x INST2	0.778**	2.42	0.02
Constant	0.775	0.73	0.47
Adj. R-squared	0.378		
N	908		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$Q = (\text{Total assets} - \text{book value of equity} + \text{market value of equity}) / \text{Total assets}$

The estimation of Eq. (12) with Tobin's Q as the dependent variable is presented in Table 6. The coefficients of Single are significantly positive, even at the 1% level. Similarly, the coefficient of Male is positive and significant at the 5% level. This indicates that both a Single CEO and Male CEO enhance the value of the firm. The result regarding Male CEO seems to indicate that the *risk-aversion* effect seems to dominate the *competency* effect in the sense that the female CEO seems to take less risk than the optimum risk level for the firm, and, therefore, a lower firm value. Also, with Tobin's Q, both the SIZE and INST seem to have nonlinear effect. This result is quite different from the one obtained with EVS as the dependent variable. The coefficient of CFS is significantly positive, even at the 1% level. This implies free cash flows have a positive impact on the firm value. Similarly, the coefficient of LTD is significantly negative at the 1% level, indicating a negative impact of leverage on firm value.

73.5 Conclusion

In this chapter, we present a brief review of nonlinear models used in the corporate finance research. We also briefly present nonlinear specification used in the empirical studies. We also present four hypotheses regarding the effects of CEO gender and marital status on a firm's R&D investment and firm values.

We also perform an empirical analysis using the US data for 2008 to test these hypotheses. Our empirical analysis indicates that there is no association between the

R&D investment and CEO gender. Similarly, there is no association between the R&D investment and CEO marital status. However, there is significant association between the R&D investment and variables like firm value, free cash flows, and firm leverage.

As to the hypothesis regarding firm value and CEO gender and marital status, we find some interesting results. Using the usual Tobin's Q to represent the firm value, we find that both single CEO and male CEO enhance the firm value. There are nonlinear relationships between firm value and each of the two control variables, firm size and institutional ownership. Finally, we find a significant positive association between the free cash flows and firm value but a significantly negative association between the firm value and leverage.

References

- Adams, R.B., and D. Ferreira. 2009. Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics* 94: 291–309.
- Aggarwal, R., L. Krigman, and K. Womack. 2002. Strategic IPO underpricing, information momentum, and lockup expiration selling. *Journal of Financial Economics* 66: 105–137.
- Ahem, K.R., and A.K. Dittmar. 2012. The changing of the boards: The impact on firm valuation of mandated female board representation. *Quarterly Journal of Economics* 127: 137–197.
- Alangar, S., C. Bathala, and R. Rao. 1999. The effect of institutional interest on the information content of dividend-change announcements. *Journal of Financial Research* 22: 429–448.
- Allen, L., and A. Cebenoyan. 1991. Bank acquisitions and ownership structure: Theory and evidence. *Journal of Banking and Finance* 15: 425–448.
- Anderson, R., and D. Reeb. 2003. Founding-family ownership and firm performance. *Journal of Finance* 58: 1301–1328.
- Anderson, R., S. Mansi, and D. Reeb. 2003. Founding family ownership and the agency cost of debt. *Journal of Financial Economics* 68: 263–285.
- Andres, P., and E. Vallelado. 2008. Corporate governance in banking: The role of board of directors. *Journal of Banking and Finance* 32: 2570–2580.
- Bae, S., D. Klein, and R. Padmaraj. 1994. Event risk bond covenants, agency costs of debt and equity, and stockholder wealth. *Financial Management* 23: 28–41.
- Bagnani, E., N. Milonas, A. Saunders, and N.G. Travlos. 1994. Managers, owners, and the pricing of risky debt: An empirical analysis. *Journal of Finance* 49: 453–477.
- Barclay, M., and C. Smith Jr. 1995. The maturity structure of corporate debt. *Journal of Finance* 50: 609–631.
- Bertrand, M., and S. Mullainathan. 2003. Enjoying the quiet life? corporate governance and managerial preferences. *Journal of Political Economy* 111: 1043–1075.
- Box, G.E.P., and D.R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 26: 211–243.
- Boyle, G., R. Carter, and R. Stover. 1998. Extraordinary antitakeover provisions and insider ownership structure: The case of converting savings and loans. *Journal of Financial and Quantitative Analysis* 33: 291–304.
- Brous, P., and O. Kini. 1994. The valuation effects of equity issues and the level of institutional ownership: Evidence from analysts' earnings forecasts. *Financial Management* 23: 33–46.
- Bugeja, Martin, Z.P. Matolcsy, and H. Spiropoulos. 2012. Is there a gender gap in CEO compensation? *Journal of Corporate Finance* 18: 849–859.
- Bustamante, M. Cecilia. 2016. How do frictions affect corporate investment? A structural approach. *Journal of Financial and Quantitative Analysis* 51: 1863–1895.

- Byrd, J., and K. Hickman. 1992. Do outside directors monitor managers? Evidence from tender offer bids. *Journal of Financial Economics* 32: 195–221.
- Byrnes, J.P., D.C. Miller, and W.D. Schafer. 1999. Gender differences in risk taking: A meta-analysis. *Psychological Bulletin* 125: 367–383.
- Chang, H.S., and C.F. Lee. 1977. Using pooled time-series and cross section data to test the firm and time effects in financial analysis. *Journal of Financial and Quantitative Analysis* 12: 457–471.
- Chen, S.S., K.W. Ho, C.F. Lee, and K. Shrestha. 2004. Nonlinear models in corporate finance research: review, critique, and extension. *Review of Quantitative Finance and Accounting* 22: 141–169.
- Chen, Jie, W.S. Leung, and K.P. Evans. 2018. Female board representation, corporate innovation and firm performance. *Journal of Empirical Finance* 48: 236–254.
- Chen, Hui, Y. Xu, and J. Yang. 2021. Systematic risk, debt maturity, and the term structure of credit spreads. *Journal of Financial Economics* 13: 770–799.
- Cho, M. 1998. Ownership structure, investment, and the corporate value: An empirical analysis. *Journal of Financial Economics* 47: 103–121.
- Christopoulos, Dimitris K., and M.A. Leon-Ledesma. 2007. A long-run non-linear approach to the fisher effect. *Journal of Money, Credit and Banking* 39: 543–559.
- Coles, J., N. Daniel, and L. Lalitha. 2008. Boards: Does one size fit all? *Journal of Financial Economics* 87: 329–356.
- Cook, D., and J. Easterwood. 1994. Poison put bonds: An analysis of their economic role. *Journal of Finance* 49: 1905–1920.
- Cook, D., R. Kieschnick, and R. Van Ness. 2006. On the marketing of IPOs. *Journal of Financial Economics* 82: 35–61.
- Dallas, L. 2002. The new managerialism and diversity on corporate boards of directors. *Tulane Law Review* 76: 1363–1405.
- Demsetz, H. 1983. The structure of ownership and the theory of the firm. *Journal of Law and Economics* 26: 375–390.
- Denis, D., D. Denis, and A. Sarin. 1977a. Agency problems, equity ownership, and corporate diversification. *Journal of Finance* 52: 135–160.
- . 1977b. Ownership structure and top executive turnover. *Journal of Financial Economics* 45: 193–221.
- Diamond, D. 1991. Debt maturity structure and liquidity risk. *Quarterly Journal of Economics* 106: 709–737.
- Dunn, P. 2010. Breaking the boardroom gender barrier: The human capital of female corporate directors. *Journal of Management and Governance* 16: 557–570.
- Eagly, A.H.L., and L. Carli. 2003. The female leadership advantage: An evaluation of the evidence. *Leadership Quarterly* 14: 807–834.
- Fama, E., and M. Jensen. 1983. Separation of ownership and control. *Journal of Law and Economics* 26: 301–325.
- Fields, L., and E. Mais. 1994. Managerial voting rights and seasoned public equity issues. *Journal of Financial and Quantitative Analysis* 29: 445–457.
- Florackis, Chrisostomos, A. Kostakis and A. Ozkan. 2009. Managerial ownership and performance. *Journal of Business Research* 62: 1350–1357.
- Fondas, N. 1997. Feminization unveiled: Management qualities in contemporary writings. *Academy of Management Review* 22: 257–282.
- Fondas, N., and S. Salsalos. 2000. A different voice in the boardroom: How the presence of women directors affects board influence over management. *Global Focus* 12: 13–22.
- Ghosh, A., and W. Ruland. 1998. Managerial ownership, the method of payment for acquisitions, and executive job retention. *Journal of Finance* 53: 785–798.
- Gilmore, R.H., and C.F. Lee. 1986. Empirical tests of granger proposition on dividend controversy. *Review of Economics and Statistics* 68: 351–355.
- Ham, Charles G., Z.R. Kaplan, and M.T. Leary. 2020. Do dividends convey information about future earnings? *Journal of Financial Economics* 136: 547–570.

- Hamadi, M., and A. Heinen. 2015. Firm performance when ownership is very concentrated: Evidence from a semiparametric panel. *Journal of Empirical Finance* 34: 172–194.
- Hambrick, D.C. 2007. Upper echelons theory: An update. *Academy of Management Review* 32: 334–343.
- Hambrick, D.C., and P.A. Mason. 1984. Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review* 9: 193–206.
- Hart, Oliver. 1983. The market mechanism as an incentive scheme. *Bell Journal of Economics* 14: 368–382.
- Hertzel, M., and R. Smith. 1993. Market discounts and shareholder gains for placing equity privately. *Journal of Finance* 48: 459–485.
- Himmelberg, C., R. Hubbard, and D. Palia. 1999. Understanding the determinants of managerial ownership and the link between ownership and performance. *Journal of Financial Economics* 53: 353–384.
- Holmstrom, Bengt. 1979. Moral hazard and observability. *Bell Journal of Economics* 10: 74–91.
- Holderness, C., R. Kroszner, and D. Sheehan. 1999. Were the good old days that good? Changes in managerial stock ownership since the great depression. *Journal of Finance* 54: 435–469.
- Jensen, M., and W. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3: 305–360.
- Jianakoplos, N., and A. Bernasek. 1998. Are women more risk averse? *Economic Inquiry* 36: 620–630.
- Joh, S. 2003. Corporate governance and firm profitability: Evidence from Korea before the economic crisis. *Journal of Financial Economics* 68: 287–322.
- Johnson, J., and P. Powell. 1994. Decision making, risk and gender: Are managers different? *British Journal of Management* 5: 123–138.
- Kim, D., D. Palia, and A. Saunders. 2010. Are initial returns and underwriting spreads in equity issues complements or substitutes? *Financial Management* 39: 1403–1423.
- Lee, C.F. 1976. Functional form and the dividend effect of the electric utility industry. *Journal of Finance* 31: 1481–1486.
- Lee, C.F., M.C. Gupta, H.Y. Chen, and A.C. Lee. 2011. Optimal payout ratio under uncertainty and the flexibility hypothesis: Theory and empirical evidence. *Journal of Corporate Finance* 17 (3): 483–501.
- Leland, H., and D. Pyle. 1977. Information asymmetries, financial structure, and financial intermediation. *Journal of Finance* 32: 371–387.
- Levin, I., M. Snyder, and D. Chapman. 1988. The interaction of experimental and situational factors and gender in a simulated risky decision-making task. *Journal of Psychology* 122: 173–181.
- Lie, E. 2000. Excess funds and agency problems: An empirical study of incremental cash distributions. *Review of Financial Studies* 13: 219–248.
- Martin, K. 1996. The method of payment in corporate acquisitions, investment opportunities, and management ownership. *Journal of Finance* 51: 1227–1246.
- Matsa, D.A., and A.R. Miller. 2013. A female style in corporate leadership? Evidence from quotas. *American Economic Journal: Applied Economics* 5: 136–169.
- Maznevski, M. 1994. Understanding our differences: Performance in decision-making groups with diverse members. *Human Relations* 47: 531–552.
- McConnell, J., and H. Servaes. 1990. Additional evidence on equity ownership and corporate value. *Journal of Financial Economics* 27: 595–612.
- McWilliams, V. 1990. Managerial share ownership and the stock price effects of antitakeover amendment proposals. *Journal of Finance* 45: 1627–1640.
- Mikkelsen, W., M. Partch, and K. Shah. 1997. Ownership and operating performance of companies that go public. *Journal of Financial Economics* 44: 281–308.
- Morck, R., A. Shleifer, and R. Vishny. 1988. Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics* 20: 293–315.
- Opler, T., L. Pinkowitz, R. Stulz, and R. Williamson. 1999. The determinants and implications of corporate cash holdings. *Journal of Financial Economics* 52: 3–46.

- Robinson, G., and K. Dechant. 1997. Building a business case for diversity. *Academy of Management Executive* 11: 21–31.
- Rosenstein, S., and J. Wyatt. 1997. Inside directors, board effectiveness, and shareholder wealth. *Journal of Financial Economics* 44: 229–250.
- Roth, G., and G. McDonald. 1999. Shareholder-management conflict and event risk covenants. *Journal of Financial Research* 22: 207–226.
- Roussanov, Nikolai, and P. Savor. 2014. Marriage and managers' attitudes to risk. *Management Science* 60: 2496–2508.
- Schooley, D., and L. Barney Jr. 1994. Using dividend policy and managerial ownership to reduce agency costs. *Journal of Financial Research* 17: 363–373.
- Setia-Atmaja, L., G. Tanewski, and M. Skully. 2009. The role of dividends, debt and board structure in the governance of family controlled firms. *Journal of Business Finance and Accounting* 36: 863–898.
- Short, H., and K. Keasey. 1999. Managerial ownership and the performance of firms: Evidence from the UK. *Journal of Corporate Finance* 5: 79–101.
- Slovin, M., M. Sushka, and K. Lai. 2000. Alternative flotation methods, adverse selection, and ownership structure: Evidence from seasoned equity issuance in the U.K. *Journal of Financial Economics* 57: 157–190.
- Stohs, M., and D. Mauer. 1996. The determinants of corporate debt maturity structure. *Journal of Business* 69: 279–312.
- Stulz, R. 1988. Managerial control of voting rights: Financing policies and the market for corporate control. *Journal of Financial Economics* 20: 25–54.
- Subrahmanyam, V., N. Rangan, and S. Rosenstein. 1997. The role of outside directors in bank acquisitions. *Financial Management* 26: 23–36.
- Sunden, A., and B. Surette. 1998. Gender differences in the allocation of assets in retirement savings plans. *American Economic Review* 88: 207–211.
- Thomas, H. 1995. Effects of firm ownership structure on hiring of M&a advisors. *Journal of Financial Services Research* 9: 159–175.
- Wood, W., D. Polek, and C. Aiken. 1985. Sex differences in group task performance. *Journal of Personality and Social Psychology* 48: 63–71.
- Wruck, K. 1989. Equity ownership concentration and firm value: Evidence from private equity financings. *Journal of Financial Economics* 23: 3–28.
- Yermack, D. 1996. Higher market valuation of companies with a small board of directors. *Journal of Financial Economics* 40: 185–211.
- Yu, H., B. Sopranzetti, and C. Lee. 2012. Multiple banking relationships, managerial ownership concentration and firm value: A simultaneous equations approach. *Quarterly Review of Economics and Finance* 52: 286–297.
- Zarembka, P. 1968. Functional form in the demand for money. *Journal of the American Statistical Association* 63: 502–511.



Three Alternative Methods for Estimating Hedge Ratios

74

Sheng-Syan Chen, Cheng-Few Lee, Fu-Lai Lin, and Keshab Shrestha

Abstract

This chapter first discusses four different theoretical models, which include minimum variance, mean-variance, expected utility, and value-at-risk method. Then we use S&P 500 data to show how three alternative estimation methods can be used to estimate hedge ratio. These three methods include OLS method, GARCH method, and cointegration and error correction method. We found that OLS method is not sufficient for estimating hedge ratio.

Keywords

Hedge ratio · Cointegration · Minimum variance · Mean-variance hedge ratio · ARCH GARCH model

S.-S. Chen (✉)

Department of Finance, College of Commerce, National Chengchi University, Taipei City, Taiwan
e-mail: sschenfn@nccu.edu.tw

C.-F. Lee

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA
e-mail: cflee@business.rutgers.edu

F.-L. Lin

Department of Finance, Dar-Yeh University, Changhua, Taiwan
e-mail: flin@mail.dyu.edu.tw

K. Shrestha

Department of Finance, School of Business, Monash University Malaysia, Bandar Sunway, Selangor, Malaysia
e-mail: keshab.shrestha@monash.edu

74.1 Introduction

A **future contract** is a standardized legal agreement between a buyer and a seller, who promise now to exchange a specified amount of money for goods or services at a future time. Of course, there is nothing really unusual about a contract made in advance of delivery. For instance, whenever something is ordered rather than purchased on the spot, a futures (or forward) contract is involved. Although the price is determined at the time of the order, the actual exchange of cash for the merchandise takes place later. For some items, the lag is a few days, while for others (such as a car) it may be months. Moreover, a futures contract imparts a legal obligation to both parties of the contract to fulfill the specifications. To guarantee fulfillment of this obligation, a “good-faith” deposit, also called **margin**, may be required from the buyer (and the seller, if he or she does not already own the product).

Future contracts are frequently used to hedge spot contracts to review the risk. Therefore, hedge ratio estimation is important for risk management. One of the best uses of derivative securities such as futures contracts is used to hedge the risk. For the last half century, both academicians and practitioners have shown great interest in the issue of hedging with futures, therefore theoretical analysis and empirical investigation for future hedge ratio are important for both academicians and practitioners. This is quite evident from the large number of articles written in this area.

The main theoretical issue in hedging involves the determination of the optimal hedge ratio. However, the optimal hedge ratio depends on the particular objective function to be optimized. In this chapter, we will carefully discuss different objective functions used to derive optimal hedge ratio. For example, one of the most widely-used hedging strategies is based on the minimization of the variance of the hedged portfolio (e.g., see Johnson 1960; Ederington 1979; and Myers and Thompson 1989). This so-called minimum-variance (MV) hedge ratio is simple to understand and estimate. However, the MV hedge ratio completely ignores the expected return of the hedged portfolio. Therefore, this strategy is inconsistent with the mean-variance tradeoff framework unless the individuals are infinitely risk-averse or the futures price follows a pure martingale process (i.e., expected futures price change is zero).

Some more realistic strategies that incorporate both the expected return and risk (variance) of the hedged portfolio have been developed (e.g., see Howard and D’Antonio 1984; Cecchetti et al. 1988; and Hsin et al. 1994). These strategies are consistent with the mean-variance framework. In fact, it can be shown that if the future price follows a pure martingale process, then the optimal mean-variance hedge ratio will be reduced to the MV hedge ratio.

Another aspect of the mean-variance-based strategies is that even though they are an improvement over the MV strategy, for them to be consistent with the expected utility maximization principle, either the utility function needs to be quadratic or the returns should be jointly normal. Unless these assumptions are valid, then the hedge ratio may not be optimal with respect to the expected utility maximization principle. Some research has solved this problem by deriving the optimal hedge ratio based on

the maximization of the expected utility (e.g., see Cecchetti et al. 1988; and Lence 1995, 1996). It should be noted that this approach requires the use of specific utility function and specific return distribution.

Attempts have been made to eliminate these specific assumptions regarding the utility function and return distributions. Some of them involve the minimization of the mean extended-Gini (MEG) coefficient, which is consistent with the concept of stochastic dominance (e.g., see Cheung et al. 1990; Kolb and Okunev 1992, 1993; Lien and Luo 1993a; Shalit 1995; and Lien and Shaffer 1999). Shalit (1995) shows that if the prices are normally distributed, then the MEG-based hedge ratio will be the same as the MV hedge ratio.

Recently, hedge ratios based on the generalized semivariance (GSV) or lower partial moments have been proposed (e.g., see De Jong et al. 1997; Lien and Tse 1998, 2000; and Chen et al. 2001). These hedge ratios are also consistent with the concept of stochastic dominance. Furthermore, these GSV-based hedge ratios have another attractive feature, whereby they measure portfolio risk by the GSV, which is consistent with the risk perceived by managers, because of its emphasis on the returns below the target return (see Crum et al. 1981; and Lien and Tse 2000). Lien and Tse (1998) show that if the futures and spot returns are jointly normally distributed, and if the futures price follows a pure martingale process, then the minimum-GSV hedge ratio will be equal to the MV hedge ratio. Finally, Hung et al. (2006) have proposed a related hedge ratio that minimizes the value-at-risk associated with the hedged portfolio when choosing hedge ratio. This hedge ratio will also be equal to MV hedge ratio if the futures price follows a pure martingale process. However, the future price in general does not follow the pure martingale process; therefore, generalized hedge ratio should be considered to improve the effectiveness of hedging decision.

It should be noted that most of the studies mentioned above (except Lence (1995, 1996)) ignore transaction costs as well as investments in other securities. Lence (1995, 1996) derives the optimal hedge ratio where transaction costs and investments in other securities are incorporated into the model. Using a CARA utility function, Lence finds that under certain circumstances, the optimal hedge ratio is zero; that is, the optimal hedging strategy is not to hedge at all.

In addition to the use of different objective functions in the derivation of the optimal hedge ratio, previous studies also differ in terms of the dynamic nature of the hedge ratio. For example, some studies assume that the hedge ratio is constant over time. Consequently, these static hedge ratios are estimated using unconditional probability distributions (e.g., see Ederington 1979; Howard and D'Antonio 1984; Benet 1992; Kolb and Okunev 1992, 1993; and Ghosh 1993). On the other hand, several studies allow the hedge ratio to change over time. In some cases, these dynamic hedge ratios are estimated using conditional distributions associated with models such as ARCH and GARCH (e.g., see Cecchetti et al. 1988; Baillie and Myers 1991; Kroner and Sultan 1993; and Sephton 1993). The GARCH based method has recently been extended by Lee and Yoder (2007) where regime-switching model is used. Alternatively, the hedge ratios can be made dynamic by considering a multi-period model where the hedge ratios are allowed to vary for different periods. This is the method used by Lien and Luo (1993b).

When it comes to estimating the hedge ratios, many different techniques are currently being employed, ranging from simple to complex ones. For example, some of them use such a simple method as the ordinary least squares (OLS) technique (e.g., see Ederington 1979; Malliaris and Urrutia 1991; and Benet 1992). However, others use more complex methods, such as the conditional heteroscedastic (ARCH or GARCH) method (e.g., see Cecchetti et al. 1988; Baillie and Myers 1991; and Sephton 1993), the random coefficient method (e.g., see Grammatikos and Saunders 1983), the cointegration method (e.g., see Ghosh 1993; Lien and Luo 1993b; and Chou et al. 1996), or the cointegration-heteroscedastic method (e.g., see Kroner and Sultan 1993). Recently, Lien and Shrestha (2007), Maghyereh et al. (2019), and Sultan et al. (2019) have suggested the use of wavelet analysis to match the data frequency with the hedging horizon. Finally, Lien and Shrestha (2010) also suggest the use of multivariate skew-normal distribution in estimating the minimum variance hedge ratio.

By now it is clear that there are several different ways of deriving and estimating hedge ratios. In this chapter, we will review these different techniques and approaches, and examine their relations. In addition, we also use S&P 500 index future data to show how three different econometrics methods can be used to estimate hedge ratios.

There are four sections in this chapter. In section 74.2, four alternative theories for deriving the static optimal hedge ratios are carefully reviewed. Section 74.3 will use three alternative estimation methods to perform an empirical study to estimate minimal hedge ratio. Finally, in Sect. 74.4 we provide a summary. Appendix A presents the monthly data of S&P 500 index and its future from January 2005 to August 2020, and Appendix B explains the applications of R language in estimating three optimal hedge ratios.

74.2 Alternative Theories Used for Deriving the Static Optimal Hedge Ratios

By combining investments in the spot market and futures market to form a portfolio, we can eliminate (or reduce) fluctuations in its value by hedging. Consider a portfolio consisting of C_s units of a long spot position and C_f units of a short futures position.¹ Following Lee et al. (2013) and Chen et al. (2003), let S_t and F_t denote the spot and futures prices at time t , respectively. Since the futures contracts are used to reduce the fluctuations in spot positions, the resulting portfolio is known as the hedged portfolio. If the hedge ratio is discussed in terms of price changes (profits), then the profit on the hedged portfolio, ΔV_H , and the hedge ratio, H , are, respectively, given by:

¹ Without loss of generality, we assume that the size of the future contract is 1.

$$\Delta V_H = C_s \Delta S_t - H C_f \Delta F_t \text{ and } H = \frac{C_f}{C_s}, \quad (1)$$

where $\Delta S_t = S_{t+1} - S_t$ and $\Delta F_t = F_{t+1} - F_t$.

The main objective of hedging is to determine and estimate the optimal hedge ratio (either h or H). It should be noted that optimal hedge ratio will depend on a particular objective function to be used. Hedge ratio can be classified into either static hedge ratio or dynamic hedge ratio.

It is important to note that in the above setup, the cash position is assumed to be fixed, and we only look for the optimum futures position. Most of the hedging literature assumes that the cash position is fixed, a setup that is suitable for financial futures. However, when we are dealing with commodity futures, the initial cash position becomes an important decision variable that is tied to the production decision, which has been discussed by Lence (1995, 1996).

Static hedge ratio means the hedge ratio remains constant over time. In general, there are nine static hedge ratios as follows: (a) Minimum-Variance Hedge Ratio, (b) Optimum Mean-Variance Hedge Ratio, (c) Sharpe Hedge Ratio, (d) Maximum Expected Utility Hedge Ratio, (e) Minimum Mean Extended-Gini Coefficient Hedge Ratio, (f) Optimum Mean-MEG Hedge Ratio, (g) Minimum Generalized Semivariance Hedge Ratio, (h) Optimum Mean-Generalized Semivariance Hedge Ratio, and (i) Minimum Value-at-Risk Hedge Ratio. In this section, we will discuss static hedge ratios (a), (b), (c), and (i) in detail. The information for the other five static hedge ratios can be found in Chen et al. (2003).

74.2.1 Minimum-Variance Hedge Ratio

The most widely used static hedge ratio is the minimum-variance (MV) hedge ratio. Johnson (1960) derives this hedge ratio by minimizing the portfolio risk, where the risk is given by the variance of changes in the value of the hedged portfolio as follows:

$$Var(\Delta V_H) = C_s^2 Var(\Delta S) + C_f^2 Var(\Delta F) - 2C_s C_f HCov(\Delta S, \Delta F).$$

The MV hedge ratio, in this case, is given by:

$$H_J^* = \frac{C_f}{C_s} = \frac{Cov(\Delta S, \Delta F)}{Var(\Delta F)}. \quad (2a)$$

Alternatively, if we use $Var(R_h)$ to represent the risk of hedged portfolio return (R_h), then the MV hedge ratio is obtained by minimizing $Var(R_h)$ which is given by:

$$Var(R_h) = Var(R_s) + h^2 Var(R_f) - 2hCov(R_s, R_f).$$

where h is the so-called hedge ratio, R_s and R_f are the so-called one-period returns on the spot and futures positions, respectively. In this case, the MV hedge ratio is given by:

$$h_f^* = \frac{Cov(R_s, R_f)}{Var(R_f)} = \rho \frac{\sigma_s}{\sigma_f}, \quad (2b)$$

where ρ is the correlation coefficient between R_s and R_f , and σ_s and σ_f are standard deviations of R_s and R_f , respectively.

The attractive features of the MV hedge ratio are that it is easy to understand and simple to compute. However, in general, the MV hedge ratio is not consistent with the mean-variance framework since it ignores the expected return on the hedged portfolio. For the MV hedge ratio to be consistent with the mean-variance framework, either the investors need to be infinitely risk-averse or the expected return on the futures contract needs to be zero.

74.2.2 Optimum Mean-Variance Hedge Ratio

Various studies have incorporated both risk and return in the derivation of the hedge ratio. For example, Hsin et al. (1994) derive the optimal hedge ratio that maximizes the following utility function:

$$\underset{C_f}{Max} \quad V(E(R_h), \sigma; A) = E(R_h) - 0.5A\sigma_h^2, \quad (3)$$

where A represents the risk aversion parameter. It is clear that this utility function incorporates both risk and return. Therefore, the hedge ratio based on this utility function would be consistent with the mean-variance framework. The optimal number of futures contract and the optimal hedge ratio are, respectively, given by:

$$h_2 = -\frac{C_f^* F}{C_s S} = -\left[\frac{E(R_f)}{A\sigma_f^2} - \rho \frac{\sigma_s}{\sigma_f} \right]. \quad (4)$$

One problem associated with this type of hedge ratio is that in order to derive the optimum hedge ratio, we need to know the individual's risk aversion parameter. Furthermore, different individuals will choose different optimal hedge ratios, depending on the values of their risk aversion parameter.

Since the MV hedge ratio is easy to understand and simple to compute, it will be interesting and useful to know under what condition the above hedge ratio would be the same as the MV hedge ratio. It can be seen from Eqs. (2b) and (4) that if $A \rightarrow \infty$ or $E(R_f) = 0$, then h_2 would be equal to the MV hedge ratio h_f^* . The first condition is simply a restatement of the infinitely risk-averse individuals. However, the second condition does not impose any condition on the risk-averseness, and this is important. It implies that even if the individuals are not infinitely risk averse, then the MV hedge ratio would be the same as the optimal mean-variance hedge ratio if the expected return on the futures contract is zero (i.e., futures prices follow a simple martingale process). Therefore, if futures prices follow a simple martingale process,

then we do not need to know the risk aversion parameter of the investor to find the optimal hedge ratio.

74.2.3 Sharpe Hedge Ratio

Another way of incorporating the portfolio return into the hedging strategy is to use the risk-return tradeoff (Sharpe measure) criteria. Howard and D'Antonio (1984) consider the optimal level of futures contracts by maximizing the ratio of the portfolio's excess return to its volatility:

$$\underset{C_f}{Max} \theta = \frac{E(R_h) - R_F}{\sigma_h}, \quad (5)$$

where C_f is the level of futures contracts. $\sigma_h^2 = Var(R_h)$ and R_F represents the risk-free interest rate.

The classic one-to-one hedge is a naïve strategy based upon a broadly defined objective of risk minimization. The strategy is naïve in the sense that a hedging coefficient of one is used regardless of past or expected correlations of spot- and futures-price changes. Working's (1953) strategy brings out the speculative aspects of hedging by analyzing changes in the basis and, accordingly, exercising discrete judgment about when to hedge and when not to hedge. The underlying objective of Working's (1953) decision rule for hedgers is one of profit maximization. Finally, Johnson (1960), in applying the mean-variance criteria of modern portfolio theory, emphasizes the risk-minimization objective but defines risk in terms of the variance of the hedged position. Although Johnson's method improves on the naïve strategy of a one-to-one hedge, however, it essentially disregards the return component associated with a particular level of risk. Rutledge (1972) uses both mean and variance information to derive hedge ratio.

Howard and D'Antonio (1984) use an objective function defined in Eq. 5 to derive the hedge ratio as defined in Eq. 6.

In deriving this hedge ratio, they use a mean-variance framework, and their strategy begins by assuming that the "agent" is out to maximize the expected return for a given level of portfolio risk. With a choice of putting money into three assets – a spot position, a futures contract, and a risk-free asset – the agent's optimal portfolio will depend on the relative risk-return characteristics of each asset. For a hedger, the optimal portfolio may contain a short futures position, a long futures position, or no futures position at all. In general, the precise futures position to be entered into will be determined by (1) the risk-free rate, (2) the expected returns and the standard deviations for the spot and futures positions, and (3) the correlation between the return on the spot position and the return on the futures.

$$\text{Hedge ratio } H = \frac{(\lambda - \rho)}{\gamma\pi(1 - \lambda\rho)} \quad (6)$$

where:

$\pi = \sigma_f / \sigma_s$ = relative variability of futures and spot returns;

$\alpha = \bar{r}_f / (\bar{r}_s - i)$ = relative excess return on futures to that of spot;

$\gamma = P_f / P_s$ = current price ratio of futures to spot;

$\lambda = \alpha / \pi = (\bar{r}_f / \sigma_f) / [(\bar{r}_s - i) / \sigma_s]$ = risk-to-excess-return relative of futures versus the spot position;

P_s, P_f = the current price per unit for the spot and futures, respectively;

ρ = simple correlation coefficient between the spot and futures returns;

σ_s = standard deviation of spot returns;

σ_f = standard deviation of futures returns;

$\bar{r}_s$ = mean return on the spot over some recent past interval;

$\bar{r}_f$ = mean return on the futures over some recent past interval; and

i = risk-free rate.

Based upon Eqs. 6 and 1, the optimal number of futures positions, C_f^* , is given by:

$$C_f^* = -C_s \frac{\left(\frac{S}{F}\right) \left(\frac{\sigma_s}{\sigma_f}\right) \left[\frac{\sigma_s}{\sigma_f} \left(\frac{E(R_f)}{E(R_s) - R_F} \right) - \rho \right]}{\left[1 - \frac{\sigma_s}{\sigma_f} \left(\frac{E(R_f)\rho}{E(R_s) - R_F} \right) \right]}. \quad (7)$$

From the optimal futures position, we can obtain the following optimal hedge ratio:

$$h_3 = - \frac{\left(\frac{\sigma_s}{\sigma_f}\right) \left[\frac{\sigma_s}{\sigma_f} \left(\frac{E(R_f)}{E(R_s) - R_F} \right) - \rho \right]}{\left[1 - \frac{\sigma_s}{\sigma_f} \left(\frac{E(R_f)\rho}{E(R_s) - R_F} \right) \right]}. \quad (8)$$

Again, if $E(R_f) = 0$, then h_3 reduces to:

$$h_3 = \left(\frac{\sigma_s}{\sigma_f} \right) \rho, \quad (9)$$

which is the same as the MV hedge ratio h_f^* .

By analyzing the properties of λ Howard and D'Antonio (1984) discern some important insights for the coordinated use of futures in a hedge portfolio. Numerically, λ expresses the relative attractiveness of investing in futures versus the spot position. When $\lambda < 1$, $\lambda = 1$, and $\lambda > 1$, the futures contract offers less, the same, and more excess return per unit of risk than the spot position, respectively. Since this analysis is being undertaken from a hedger's point of view, it is assumed $\lambda < 1$. An assumption that $\lambda > 1$ would inappropriately imply that theoretically it is possible to hedge the futures position with the spot asset. In addition, they also consider the effect of the relationship of λ to ρ on the optimal hedge ratio, H . Other detailed analysis can be found in Table 4 of Howard and D'Antonio's (1984) paper.

74.2.4 Minimum Value-at-Risk Hedge Ratio

Hung et al. (2006) suggest a new hedge ratio that minimizes the Value-at-Risk of the hedged portfolio. Specifically, the hedge ratio h is derived by minimizing the following Value-at-Risk of the hedged portfolio over a given time period τ :

$$VaR(R_h) = Z_\alpha \sigma_h \sqrt{\tau} - E[R_h] \tau \quad (10)$$

The resulting optimal hedge ratio, which Hung et al. (2006) refer to as zero-VaR hedge ratio, is given by

$$h^{VaR} = \rho \frac{\sigma_s}{\sigma_f} - E[R_f] \frac{\sigma_s}{\sigma_f} \sqrt{\frac{1 - \rho^2}{Z_\alpha^2 \sigma_f^2 - E[R_f]^2}} \quad (11)$$

It is clear that, if the futures price follows martingale process, the zero-VaR hedge ratio would be the same as the MV hedge ratio.

74.3 Applications of OLS, GARCH, and CECM Models to Estimate Optimal Hedge Ratio

In this section, we will discuss different approaches to estimating the optimum hedge ratios and apply these methods in practice to estimate the optimal stock index futures hedge ratio for S&P 500 index. We first introduce how to apply the ordinary least square (OLS) technique to estimating the optimum hedge ratios and discuss its restrictions. Secondly, considering the heteroscedastic nature of the error term, generalized autoregressive conditional heteroskedasticity (GARCH) technique is introduced to estimating the hedge ratio. In addition, we describe cointegration and error correction method to take into account the possibility of long-run equilibrium between spot and futures prices. Finally, we provide an empirical study using OLS, GARCH, and CECM models to estimate the optimal stock index futures hedge ratio for S&P 500 index through R language.

74.3.1 OLS Method

The conventional approach to estimating the MV hedge ratio involves the regression of the changes in spot prices on the changes in futures price using the OLS technique (e.g., see Junkus and Lee 1985). Specifically, the regression equation can be written as:

$$\Delta S_t = a_0 + a_1 \Delta F_t + e_t, \quad (12)$$

where the estimate of the MV hedge ratio, H_j , is given by a_1 . The OLS technique is quite robust and simple to use. However, for the OLS technique to be valid and

efficient, assumptions associated with the OLS regression must be satisfied. One case where the assumptions are not completely satisfied is that the error term in the regression is heteroscedastic. This situation will be discussed later.

Another problem with the OLS method, as pointed out by Myers and Thompson (1989), is the fact that it uses unconditional sample moments instead of conditional sample moments, which use currently available information. They suggest the use of the conditional covariance and conditional variance in Eq. (2a). In this case, the conditional version of the optimal hedge ratio (Eq. (2a)) will take the following form:

$$H_J^* = \frac{C_f}{C_s} = \frac{\text{Cov}(\Delta S, \Delta F) \mid \Omega_{t-1}}{\text{Var}(\Delta F) \mid \Omega_{t-1}}. \quad (13)$$

Suppose that the current information (Ω_{t-1}) includes a vector of variables (X_{t-1}) and the spot and futures price changes are generated by the following equilibrium model:

$$\Delta S_t = X_{t-1}\alpha + u_t,$$

$$\Delta F_t = X_{t-1}\beta + v_t.$$

In this case, the maximum likelihood estimator of the MV hedge ratio is given by (see Myers and Thompson (1989)):

$$\hat{h} \mid X_{t-1} = \frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}, \quad (14)$$

where $\hat{\sigma}_{uv}$ is the sample covariance between the residuals u_t and v_t , and $\hat{\sigma}_v^2$ is the sample variance of the residual v_t . In general, the OLS estimator obtained from Eq. (12) would be different from the one given by Eq. (14). For the two estimators to be the same, the spot and futures prices must be generated by the following model:

$$\Delta S_t = \alpha_0 + u_t, \quad \Delta F_t = \beta_0 + v_t.$$

In other words, if the spot and futures prices follow a random walk, then with or without drift, the two estimators will be the same. Otherwise, the hedge ratio estimated from the OLS regression (12) will not be optimal.

74.3.2 GARCH Method

Ever since the development of GARCH models, the OLS method of estimating the hedge ratio has been generalized to consider the heteroscedastic nature of the error term in Eq. (12). In this case, rather than using the unconditional sample variance and covariance, the conditional variance and covariance from the GARCH model are

used in the estimation of the hedge ratio. As mentioned above, such a technique allows an update of the hedge ratio over the hedging period.

Consider the following bivariate GARCH model (see Cecchetti et al. 1988; and Baillie and Myers 1991):

$$\begin{bmatrix} \Delta S_t \\ \Delta F_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \text{ and } e_t \mid \Omega_{t-1} \sim N(0, H_t)$$

where Ω_{t-1} is all the available information up to time $t-1$. N denotes a bivariate normal distribution, and H_t is a time-varying 2×2 positive-definite conditional variance-covariance matrix as follows.

$$H_t = \begin{bmatrix} H_{11,t} & H_{12,t} \\ H_{12,t} & H_{22,t} \end{bmatrix}$$

The conditional MV hedge ratio at time t is given by $h_{t-1} = H_{12,t}/H_{22,t}$. This model allows the hedge ratio to change over time, resulting in a series of hedge ratios instead of a single hedge ratio for the entire hedging horizon.

In practice, it is necessary to model the conditional variance-covariance matrix and ensure the positive definiteness of H_t . Different models for H_t have been proposed in the literature, for example, Bollerslev (1988), Bollerslev et al. (1988) and Engle and Kroner (1995). Bollerslev (1988) built up a VECH-GARCH (1,1) model assuming that H_t is a linear combination of its own squared residual e'_{t-1} and past conditional variance-covariance variance H_{t-1} as follows.

$$\text{vec}(H_t) = C + A \text{vec}(e_{t-1}e'_{t-1}) + B \text{vec}(H_{t-1}).$$

where vec is the operator that stacks a matrix as a column vector. A and B are both a 3×3 parameter matrices.

$$\begin{aligned} \text{vec}(H_t) &= \begin{bmatrix} H_{11,t} \\ H_{12,t} \\ H_{22,t} \end{bmatrix} \\ &= \begin{bmatrix} c_{01,t} \\ c_{02,t} \\ c_{03,t} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{3,t-1}^2 \end{bmatrix} + \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \\ &\quad \times \begin{bmatrix} H_{11,t-1} \\ H_{12,t-1} \\ H_{13,t-1} \end{bmatrix} \end{aligned}$$

To simplify calculation, Bollerslev et al. (1988) suggest a diagonal VECH-GARCH (1,1) model that parameter matrices A and B are restricted to be diagonal.

$$\begin{aligned}
\text{vec}(H_t) &= \begin{bmatrix} H_{11,t} \\ H_{12,t} \\ H_{22,t} \end{bmatrix} \\
&= \begin{bmatrix} c_{11}^* \\ c_{12}^* \\ c_{22}^* \end{bmatrix} + \begin{bmatrix} a_{11}^* & 0 & 0 \\ 0 & a_{12}^* & 0 \\ 0 & 0 & a_{22}^* \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} g_{11}^* & 0 & 0 \\ 0 & g_{12}^* & 0 \\ 0 & 0 & g_{22}^* \end{bmatrix} \\
&\quad \times \begin{bmatrix} H_{11,t-1} \\ H_{12,t-1} \\ H_{22,t-1} \end{bmatrix}
\end{aligned}$$

74.3.3 Cointegration and Error Correction Method (CECM)

The techniques described so far do not take into consideration the possibility that spot price and futures price series could be non-stationary. If these series have unit roots, then this will raise a different issue. If the two series are cointegrated as defined by Engle and Granger (1987), then the regression Eq. (12) will be mis-specified and an error-correction term must be included in the equation. Since the arbitrage condition ties the spot and futures prices, they cannot drift far apart in the long run. Therefore, if both series follow a random walk, then we expect the two series to be cointegrated in which case we need to estimate the error correction model. This calls for the use of the cointegration analysis. Therefore, the cointegration and error correction method should be used to estimate the hedge ratio.

The cointegration analysis involves two steps. First, each series must be tested for a unit root (e.g., see Dickey and Fuller 1981; and Phillips and Perron 1988). Second, if both series are found to have a single unit root, then the cointegration test must be performed (e.g., see Engle and Granger 1987; Johansen and Juselius 1990; and Osterwald-Lenum 1992).

If the spot price and futures price series are found to be cointegrated, then the hedge ratio can be estimated in two steps (see Ghosh 1993; and Chou et al. 1996). The first step involves the estimation of the following cointegrating regression:

$$S_t = a + bF_t + u_t. \quad (15)$$

The second step involves the estimation of the following error correction model:

$$\Delta S_t = \rho u_{t-1} + \beta \Delta F_t + \sum_{i=1}^m \delta_i \Delta F_{t-i} + \sum_{j=1}^n \theta_j \Delta S_{t-j} + e_j, \quad (16)$$

where u_t is the residual series from the cointegrating regression. The estimate of the hedge ratio is given by the estimate of β . Some researchers (e.g., see Lien and Luo

1993b) assume that the long-run cointegrating relationship is $(S_t - F_t)$, and estimate the following error correction model:

$$\Delta S_t = \rho(S_{t-1} - F_{t-1}) + \beta \Delta F_t + \sum_{i=1}^m \delta_i \Delta F_{t-i} + \sum_{j=1}^n \theta_j \Delta S_{t-j} + e_j. \quad (17)$$

Alternatively, Chou et al. (1996) suggest the estimation of the error correction model as follows:

$$\Delta S_t = \alpha \hat{u}_{t-1} + \beta \Delta F_t + \sum_{i=1}^m \delta_i \Delta F_{t-i} + \sum_{j=1}^n \theta_j \Delta S_{t-j} + e_j, \quad (18)$$

where $\hat{u}_{t-1} = S_{t-1} - (a + bF_{t-1})$; that is, the series $\hat{u}_t$ is the estimated residual series from Eq. (15). The hedge ratio is given by β in Eq. (17).

Kroner and Sultan (1993) combine the error-correction model with the GARCH model considered by Cecchetti et al. (1988) and Baillie and Myers (1991) in order to estimate the optimum hedge ratio. Specifically, they use the following model:

$$\begin{bmatrix} \Delta \log_e(S_t) \\ \Delta \log_e(F_t) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \alpha_s(\log_e(S_{t-1}) - \log_e(F_{t-1})) \\ \alpha_f(\log_e(S_{t-1}) - \log_e(F_{t-1})) \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}, \quad (19)$$

where the error processes follow a GARCH process. As before, the hedge ratio at time $(t - 1)$ is given by $h_{t-1} = H_{12,t}/H_{22,t}$.

74.3.4 Empirical Studies

Here, we apply OLS, GARCH, and CECM models in practice to estimate optimal hedge ratios through R language. Monthly data for S&P 500 index and its futures were collected from Datastream database, the sample consisted of 188 observations from January 31, 2005, to August 31, 2020. Summary statistics of this data are presented in Table 1.

First, in Eq. (12), we use the OLS method by regressing the changes in spot prices on the changes in futures prices to estimate the optimal hedge ratio. The estimate of

Table 1 Summary statistics

	Min	Max	Mean	Std Dev	Skewness	Kurtosis
SPOT	735.09	3500.31	1781.57	656.57	0.66	2.30
FUTURES	734.20	3498.90	1780.09	656.05	0.66	2.31
ΔS_t	-369.63	327.84	12.17	79.23	-0.76	7.92
ΔF_t	-381.40	332.70	12.16	80.34	-0.77	8.08
Observations	188					

Table 2 Hedge ratio coefficient using the conventional regression model

Variable	Estimate	Std. Error	t-ratio	p-value
Intercept	0.1984	0.2729	0.73	0.4680
ΔF_t	0.9851	0.0034	292.53	<0.0001

Table 3 Hedge ratio coefficient using the conventional regression model with heteroscedastic errors

Variable	Estimate	Std. Error	t-ratio	p-value
Intercept	0.0490	0.0144	3.41	0.0007
ΔF_t	0.9994	0.0008	1179.59	<0.0001
e_{t-1}	-0.9873	0.0109	-90.29	<0.0001
e_{t-2}	-0.9959	0.0145	-68.83	<0.0001
ω	0.0167	0.0098	1.71	0.0866
ϵ_{t-1}^2	0.3135	0.0543	5.78	<0.0001
h_{t-1}	0.6855	0.0530	12.94	<0.0001

hedge ratio obtained from the OLS technique are reported in Table 2. As shown in Table 2, we can see that the hedge ratio of S&P 500 index is significantly different from zero, at a 1% significance level. Moreover, the estimated hedge ratio, denoted by the coefficient of ΔF_t , is generally less than unity.

Secondly, we apply a conventional regression model with heteroscedastic error terms to estimate the hedge ratio. Here, an AR(2)-GARCH(1,1) model for the changes in spot prices regressed on the changes in futures prices is specified as follow,

$$\Delta S_t = a_0 + a_1 \Delta F_t + e_t, e_t = \epsilon_t - \varphi_1 e_{t-1} - \varphi_2 e_{t-2}$$

$$\epsilon_t = \sqrt{h_t} \epsilon_t, h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}$$

where $\epsilon_t \sim N(0, 1)$. The estimated result of AR(2)-GARCH(1,1) model is shown in Table 3. The coefficient estimates of AR(2)-GARCH(1,1) model, as shown in Table 3, are all significantly different from zero, at a 1% significance level. This finding suggests that the importance of capturing the heteroscedastic error structures in conventional regression model. In addition, the hedge ratio of conventional regression with AR(2)-GARCH(1,1) model is higher than the OLS hedge ratio for S&P 500 futures contract.

Next, we will apply the CECM model to estimate optimal hedge ratio. Here, standard-augmented Dickey-Fuller (ADF) unit roots and Phillips and Ouliaris (1990) residual cointegration tests are performed, and the optimal hedge ratios estimated by error correction model (ECM) will be presented. Here, we apply the augmented Dickey-Fuller (ADF) regression to test for the presence of unit roots. The ADF test statistics, as shown in Panel A of Table 4, indicates that the null hypothesis of a unit root cannot be rejected for the levels of the variables. Using differenced data, the computed ADF test statistics shown in Panel B of Table 4 suggested that the null hypothesis is rejected, at the 1% significance level. As differencing one

Table 4 Unit roots and residual cointegration tests results

Variable	ADF statistics	lag parameter	p-value
<i>Panel A. Level data</i>			
Spot	−1.3353	1	0.8542
Futures	−1.3458	1	0.8498
<i>Panel B. First-order differenced data</i>			
Spot	−10.104	1	<0.01
Futures	−10.150	1	<0.01
<i>Panel C. Phillips-Ouliaris Cointegration Test</i>			
Phillips-Ouliaris demeaned	−60.783	1	<0.01

Table 5 Error correction estimates of hedge ratio coefficient

Variable	Estimate	Std. Error	t-ratio	p-value
ΔF_t	0.9892	0.0031	316.60	<0.001
$\hat{u}_{t-1}$	−0.3423	0.0571	−5.99	<0.001

produces stationarity, we may conclude that each series is integrated of order one, $I(1)$, process which is necessary for testing the existence of cointegration. We then apply Phillips and Ouliaris (1990) residual cointegration test to examine the presence of cointegration. The result of Phillips-Ouliaris cointegration test shown is reported in Panel C of Table 4. The null hypothesis of the Phillips-Ouliaris cointegration test is that there is no cointegration present. The result of Phillips-Ouliaris cointegration test indicates the null hypothesis of no cointegration is rejected, at the 1% significance level. This suggests that the spot S&P 500 index is cointegrated with the S&P 500 index futures.

Finally, we apply the ECM model in term of Eq. (16) to estimate optimal hedge ratio. Table 5 shows that the coefficient on the error-correction term, $\hat{u}_{t-1}$, is significantly different from zero, at a 1% significance level. This suggests that the importance of estimating error correction model, and in particular the long-run equilibrium error term cannot be ignored in conventional regression model. In addition, the ECM hedge ratio is higher than the conventional OLS hedge ratio for S&P 500 futures contract. This finding is consistent with the results in Lien (1996, 2004) who argued that the MV hedge ratio will be smaller if the cointegration relationship is not considered.

74.4 Conclusion

In this chapter, we have reviewed four approaches for deriving the optimal hedge ratio. These approaches can be divided into the minimum-variance approach, the mean-variance based approach, and the expected utility maximizing approach. All these approaches will lead to the same hedge ratio as the conventional minimum-variance (MV) hedge ratio, if the futures price follows a pure martingale process and if the futures and spot prices are jointly normal. However, if these conditions do not

hold, then the hedge ratios based on the various approaches will be different, therefore the user should carefully examine what kind of assumption they will use.

Since the statistical properties of the MV hedge ratio are well known, it is easy to estimate and perform the statistical hypothesis test. For example, we can test whether the optimal MV hedge ratio is the same as the naïve hedge ratio. Since the MV hedge ratio ignores the expected return, it will not be consistent with the mean-variance analysis unless the futures price follows a pure martingale process. In general, if the martingale and normality conditions do not hold, then the MV hedge ratio will not be consistent with the expected utility maximization principle in terms of mean and variance.

In order to make the hedge ratio consistent with the expected utility maximization principle, we can derive the optimal hedge ratio by maximizing the expected utility in terms of mean and variance. However, to implement such an approach, we need to assume a specific utility function such as quadratic utility function, and we need to make an assumption regarding the return distribution. Therefore, different utility functions will lead to different optimal hedge ratios.

It should be noted that the derivation of the optimal hedge ratio discussed in this chapter does not incorporate transaction costs. Furthermore, these derivations do not allow investments in securities other than the spot and corresponding futures contracts. As shown by Lence (1995), once we relax these conventional assumptions, the resulting optimal hedge ratio can be quite different from the ones obtained under the conventional assumptions. Lence's (1995) results are based on a specific utility function and some other assumption regarding the return distributions.

In this chapter, we have also reviewed various ways of estimating the optimum hedge ratio. As far as the estimation of the conventional MV hedge ratio is concerned, there are a large number of methods that have been proposed in the literature. These methods range from a simple regression method to complex cointegrated heteroscedastic methods with regime-switching, and some of the estimation methods include a kernel density function method as well as an empirical distribution method.

In this chapter, we have also discussed about the relationship between the optimal MV hedge ratio and the hedging horizon. We feel that this relationship has not been fully explored and can be further developed in the future. For example, we would like to know if the optimal hedge ratio approaches the naïve hedge ratio when the hedging horizon becomes longer.

In this chapter, we used S&P 500 index futures from January 2005 to August 2020 in terms of three estimation methods to estimate hedge ratio. These three methods are OLS regression method, GARCH method, and CECM method. We also compared hedge ratio estimates from these three methods in some detail. The R programs used to estimate this empirical result can be found in Appendix B.

For those who are interested in research in this area, we would like to finally point out that one requires a good understanding of financial economic theories and econometric methodologies. In addition, a good background in data analysis and computer programming would also be helpful.

Appendix A. Monthly Data of S&P500 Index and Its Futures (January 2005 – August 2020)

Date	SPOT	FUTURES	C_spot	C_futures
1/31/2005	1181.27	1181.7	−30.65	−32
2/28/2005	1203.6	1204.1	22.33	22.4
3/31/2005	1180.59	1183.9	−23.01	−20.2
4/29/2005	1156.85	1158.5	−23.74	−25.4
5/31/2005	1191.5	1192.3	34.65	33.8
6/30/2005	1191.33	1195.5	−0.17	3.2
7/29/2005	1234.18	1236.8	42.85	41.3
8/31/2005	1220.33	1221.4	−13.85	−15.4
9/30/2005	1228.81	1234.3	8.48	12.9
10/31/2005	1207.01	1209.8	−21.8	−24.5
11/30/2005	1249.48	1251.1	42.47	41.3
12/30/2005	1248.29	1254.8	−1.19	3.7
1/31/2006	1280.08	1283.6	31.79	28.8
2/28/2006	1280.66	1282.4	0.58	−1.2
3/31/2006	1294.83	1303.3	14.17	20.9
4/28/2006	1310.61	1315.9	15.78	12.6
5/31/2006	1270.09	1272.1	−40.52	−43.8
6/30/2006	1270.2	1279.4	0.11	7.3
7/31/2006	1276.66	1281.8	6.46	2.4
8/31/2006	1303.82	1305.6	27.16	23.8
9/29/2006	1335.85	1345.4	32.03	39.8
10/31/2006	1377.94	1383.2	42.09	37.8
11/30/2006	1400.63	1402.9	22.69	19.7
12/29/2006	1418.3	1428.4	17.67	25.5
1/31/2007	1438.24	1443	19.94	14.6
2/28/2007	1406.82	1408.9	−31.42	−34.1
3/30/2007	1420.86	1431.2	14.04	22.3
4/30/2007	1482.37	1488.4	61.51	57.2
5/31/2007	1530.62	1532.9	48.25	44.5
6/29/2007	1503.35	1515.4	−27.27	−17.5
7/31/2007	1455.27	1461.9	−48.08	−53.5
8/31/2007	1473.99	1476.7	18.72	14.8
9/28/2007	1526.75	1538.1	52.76	61.4
10/31/2007	1549.38	1554.9	22.63	16.8
11/30/2007	1481.14	1483.7	−68.24	−71.2
12/31/2007	1468.35	1477.2	−12.79	−6.5
1/31/2008	1378.55	1379.6	−89.8	−97.6
2/29/2008	1330.63	1331.3	−47.92	−48.3
3/31/2008	1322.7	1324	−7.93	−7.3
4/30/2008	1385.59	1386	62.89	62
5/30/2008	1400.38	1400.6	14.79	14.6

(continued)

Date	SPOT	FUTURES	C_spot	C_futures
6/30/2008	1280	1281.1	-120.38	-119.5
7/31/2008	1267.38	1267.1	-12.62	-14
8/29/2008	1282.83	1282.6	15.45	15.5
9/30/2008	1166.36	1169	-116.47	-113.6
10/31/2008	968.75	967.3	-197.61	-201.7
11/28/2008	896.24	895.3	-72.51	-72
12/31/2008	903.25	900.1	7.01	4.8
1/30/2009	825.88	822.5	-77.37	-77.6
2/27/2009	735.09	734.2	-90.79	-88.3
3/31/2009	797.87	794.8	62.78	60.6
4/30/2009	872.81	870	74.94	75.2
5/29/2009	919.14	918.1	46.33	48.1
6/30/2009	919.32	915.5	0.18	-2.6
7/31/2009	987.48	984.4	68.16	68.9
8/31/2009	1020.62	1019.7	33.14	35.3
9/30/2009	1057.08	1052.9	36.46	33.2
10/30/2009	1036.19	1033	-20.89	-19.9
11/30/2009	1095.63	1094.8	59.44	61.8
12/31/2009	1115.1	1110.7	19.47	15.9
1/29/2010	1073.87	1070.4	-41.23	-40.3
2/26/2010	1104.49	1103.4	30.62	33
3/31/2010	1169.43	1165.2	64.94	61.8
4/30/2010	1186.69	1183.4	17.26	18.2
5/31/2010	1089.41	1088.5	-97.28	-94.9
6/30/2010	1030.71	1026.6	-58.7	-61.9
7/30/2010	1101.6	1098.3	70.89	71.7
8/31/2010	1049.33	1048.3	-52.27	-50
9/30/2010	1141.2	1136.7	91.87	88.4
10/29/2010	1183.26	1179.7	42.06	43
11/30/2010	1180.55	1179.6	-2.71	-0.1
12/31/2010	1257.64	1253	77.09	73.4
1/31/2011	1286.12	1282.4	28.48	29.4
2/28/2011	1327.22	1326.1	41.1	43.7
3/31/2011	1325.83	1321	-1.39	-5.1
4/29/2011	1363.61	1359.7	37.78	38.7
5/31/2011	1345.2	1343.9	-18.41	-15.8
6/30/2011	1320.64	1315.5	-24.56	-28.4
7/29/2011	1292.28	1288.4	-28.36	-27.1
8/31/2011	1218.89	1217.7	-73.39	-70.7
9/30/2011	1131.42	1126	-87.47	-91.7
10/31/2011	1253.3	1249.3	121.88	123.3
11/30/2011	1246.96	1246	-6.34	-3.3
12/30/2011	1257.6	1252.6	10.64	6.6
1/31/2012	1312.41	1308.2	54.81	55.6

(continued)

Date	SPOT	FUTURES	C_spot	C_futures
2/29/2012	1365.68	1364.4	53.27	56.2
3/30/2012	1408.47	1403.2	42.79	38.8
4/30/2012	1397.91	1393.6	-10.56	-9.6
5/31/2012	1310.33	1309.2	-87.58	-84.4
6/29/2012	1362.16	1356.4	51.83	47.2
7/31/2012	1379.32	1374.6	17.16	18.2
8/31/2012	1406.58	1405.1	27.26	30.5
9/28/2012	1440.67	1434.2	34.09	29.1
10/31/2012	1412.16	1406.8	-28.51	-27.4
11/30/2012	1416.18	1414.4	4.02	7.6
12/31/2012	1426.19	1420.1	10.01	5.7
1/31/2013	1498.11	1493.3	71.92	73.2
2/28/2013	1514.68	1513.3	16.57	20
3/29/2013	1569.19	1562.7	54.51	49.4
4/30/2013	1597.57	1592.2	28.38	29.5
5/31/2013	1630.74	1629	33.17	36.8
6/28/2013	1606.28	1599.3	-24.46	-29.7
7/31/2013	1685.73	1680.5	79.45	81.2
8/30/2013	1632.97	1631.3	-52.76	-49.2
9/30/2013	1681.55	1674.3	48.58	43
10/31/2013	1756.54	1751	74.99	76.7
11/29/2013	1805.81	1804.1	49.27	53.1
12/31/2013	1848.36	1841.1	42.55	37
1/31/2014	1782.59	1776.6	-65.77	-64.5
2/28/2014	1859.45	1857.6	76.86	81
3/31/2014	1872.34	1864.6	12.89	7
4/30/2014	1883.95	1877.9	11.61	13.3
5/30/2014	1923.57	1921.5	39.62	43.6
6/30/2014	1960.23	1952.4	36.66	30.9
7/31/2014	1930.67	1924.8	-29.56	-27.6
8/29/2014	2003.37	2001.4	72.7	76.6
9/30/2014	1972.29	1965.5	-31.08	-35.9
10/31/2014	2018.05	2011.4	45.76	45.9
11/28/2014	2067.56	2066.3	49.51	54.9
12/31/2014	2058.9	2052.4	-8.66	-13.9
1/30/2015	1994.99	1988.4	-63.91	-64
2/27/2015	2104.5	2102.8	109.51	114.4
3/31/2015	2067.89	2060.8	-36.61	-42
4/30/2015	2085.51	2078.9	17.62	18.1
5/29/2015	2107.39	2106	21.88	27.1
6/30/2015	2063.11	2054.4	-44.28	-51.6
7/31/2015	2103.84	2098.4	40.73	44
8/31/2015	1972.18	1969.2	-131.66	-129.2
9/30/2015	1920.03	1908.7	-52.15	-60.5

(continued)

Date	SPOT	FUTURES	C_spot	C_futures
10/30/2015	2079.36	2073.7	159.33	165
11/30/2015	2080.41	2079.8	1.05	6.1
12/31/2015	2043.94	2035.4	-36.47	-44.4
1/29/2016	1940.24	1930.1	-103.7	-105.3
2/29/2016	1932.23	1929.5	-8.01	-0.6
3/31/2016	2059.74	2051.5	127.51	122
4/29/2016	2065.3	2059.1	5.56	7.6
5/31/2016	2096.96	2094.9	31.66	35.8
6/30/2016	2098.86	2090.2	1.9	-4.7
7/29/2016	2173.6	2168.2	74.74	78
8/31/2016	2170.95	2169.5	-2.65	1.3
9/30/2016	2168.27	2160.4	-2.68	-9.1
10/31/2016	2126.15	2120.1	-42.12	-40.3
11/30/2016	2198.81	2198.8	72.66	78.7
12/30/2016	2238.83	2236.2	40.02	37.4
1/31/2017	2278.87	2274.5	40.04	38.3
2/28/2017	2363.64	2362.8	84.77	88.3
3/31/2017	2362.72	2359.2	-0.92	-3.6
4/28/2017	2384.2	2380.5	21.48	21.3
5/31/2017	2411.8	2411.1	27.6	30.6
6/30/2017	2423.41	2420.9	11.61	9.8
7/31/2017	2470.3	2468	46.89	47.1
8/31/2017	2471.65	2470.1	1.35	2.1
9/29/2017	2519.36	2516.1	47.71	46
10/31/2017	2575.26	2572.7	55.9	56.6
11/30/2017	2647.58	2647.9	72.32	75.2
12/29/2017	2673.61	2676	26.03	28.1
1/31/2018	2823.81	2825.8	150.2	149.8
2/28/2018	2713.83	2714.4	-109.98	-111.4
3/30/2018	2640.87	2643	-72.96	-71.4
4/30/2018	2648.05	2647	7.18	4
5/31/2018	2705.27	2705.5	57.22	58.5
6/29/2018	2718.37	2721.6	13.1	16.1
7/31/2018	2816.29	2817.1	97.92	95.5
8/31/2018	2901.52	2902.1	85.23	85
9/28/2018	2913.98	2919	12.46	16.9
10/31/2018	2711.74	2711.1	-202.24	-207.9
11/30/2018	2760.17	2758.3	48.43	47.2
12/31/2018	2506.85	2505.2	-253.32	-253.1
1/31/2019	2704.1	2704.5	197.25	199.3
2/28/2019	2784.49	2784.7	80.39	80.2
3/29/2019	2834.4	2837.8	49.91	53.1
4/30/2019	2945.83	2948.5	111.43	110.7
5/31/2019	2752.06	2752.6	-193.77	-195.9
6/28/2019	2941.76	2944.2	189.7	191.6

(continued)

Date	SPOT	FUTURES	C_spot	C_futures
7/31/2019	2980.38	2982.3	38.62	38.1
8/30/2019	2926.46	2924.8	-53.92	-57.5
9/30/2019	2976.74	2978.5	50.28	53.7
10/31/2019	3037.56	3035.8	60.82	57.3
11/29/2019	3140.98	3143.7	103.42	107.9
12/31/2019	3230.78	3231.1	89.8	87.4
1/31/2020	3225.52	3224	-5.26	-7.1
2/28/2020	2954.22	2951.1	-271.3	-272.9
3/31/2020	2584.59	2569.7	-369.63	-381.4
4/30/2020	2912.43	2902.4	327.84	332.7
5/29/2020	3044.31	3042	131.88	139.6
6/30/2020	3100.29	3090.2	55.98	48.2
7/31/2020	3271.12	3263.5	170.83	173.3
8/31/2020	3500.31	3498.9	229.19	235.4

Appendix B. Applications of R Language in Estimating Three Different Optimal Hedge Ratios

In this appendix, we show the estimation procedure on how to apply OLS, GARCH, and CECM models to estimate optimal hedge ratios through R language. R language is a high-level computer language that is designed for statistics and graphics. Compared to alternatives, SAS, Matlab or Stata, R is completely free. Another benefit is that it is open source. Users could head to <http://cran.r-project.org/> to download and install R language. Based upon monthly S&P 500 index and its futures as presented in Appendix A, the estimation procedures of applying R language to estimate hedge ratio are provided as follows.

First, we use OLS method in term of Eq. (12) to estimate minimum variance hedge ratio. By using linear model (lm) function in R language, we obtain the following program code.

```
SP500= read.csv(file="SP500.csv")
OLS.fit <- lm(C_spot~C_futures, data=SP500)
summary(OLS.fit)
```

Next, we apply a conventional regression model with an AR(2)-GARCH(1,1) error terms to estimate minimum variance hedge ratio. By using rugarch package in R language, we obtain the following program.

```
library(rugarch)
fit.spec <- ugarchspec(
  variance.model = list(model = "sGARCH",
    garchOrder = c(1, 1)),
  mean.model = list(armaOrder = c(2, 0), include.mean = TRUE,
    external.regressors= cbind(SP500$C_futures)),
  distribution.model = "norm")
```

```
GARCH.fit <- ugarchfit(data = cbind(SP500$C_spot),
  spec = fit.spec)
GARCH.fit
```

Third, we apply the ECM model to estimate minimum variance hedge ratio. We begin by applying an augmented Dickey-Fuller (ADF) test for the presence of unit roots. The Phillips and Ouliaris (1990) residual cointegration test is applied to examine the presence of cointegration. Finally, the minimum variance hedge ratio is estimated by the error correction model. By using *tseries* package in R language, we obtain the following program.

```
library(tseries)

# Augmented Dickey-Fuller Test
# Level data

adf.test(SP500$SPOT, k = 1)
adf.test(SP500$FUTURES, k = 1)

# First-order differenced data

adf.test(diff(SP500$SPOT), k = 1)
adf.test(diff(SP500$FUTURES), k = 1)

# Phillips and Ouliaris (1990) residual cointegration test

po.test(cbind(SP500$FUTURES, SP500$SPOT))

# Engle-Granger two-step procedure
## 1.Estimate cointegrating relationship

reg <- lm(SPOT~FUTURES, data=SP500)

## 2. Compute error term

Resid <- reg$resid

# Estimate optimal hedge ratio using the error correction model

n=length(resid)
ECM.fit <-lm(diff(SPOT) ~ -1 + diff(FUTURES) + Resid[1:n-1],
data=SP500)
summary(ECM.fit)
```

Finally, we apply a multivariate GARCH(1,1)-BEKK model to estimate dynamic minimum variance hedge ratio.

```
library(mgarchBEKK)
estimated=BEKK(as.matrix(cbind(SP500$C_futures,SP500$C_spot)),
  order = c(1, 1))
estimated$est.params
estimated$asy.se.coef
```

References

- Baillie, R.T., and R.J. Myers. 1991. Bivariate Garch estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics* 6: 109–124.
- Benet, B.A. 1992. Hedge period length and ex-ante futures hedging effectiveness: The case of foreign-exchange risk cross hedges. *Journal of Futures Markets* 12: 163–175.
- Cecchetti, S.G., R.E. Cumby, and S. Figlewski. 1988. Estimation of the optimal futures hedge. *Review of Economics and Statistics* 70: 623–630.
- Chen, S.S., C.F. Lee, and K. Shrestha. 2001. On a mean-generalized semivariance approach to determining the hedge ratio. *Journal of Futures Markets* 21: 581–598.
- . 2003. Futures hedge ratios: A review. *The Quarterly Review of Economics and Finance* 43 (3): 433–465.
- Cheung, C.S., C.C.Y. Kwan, and P.C.Y. Yip. 1990. The hedging effectiveness of options and futures: A mean-Gini approach. *Journal of Futures Markets* 10: 61–74.
- Chou, W.L., K.K. Fan, and C.F. Lee. 1996. Hedging with the Nikkei index futures: The conventional model versus the error correction model. *Quarterly Review of Economics and Finance* 36: 495–505.
- Crum, R.L., D.L. Laughhunn, and J.W. Payne. 1981. Risk-seeking behavior and its implications for financial models. *Financial Management* 10: 20–27.
- De Jong, A., F. De Roon, and C. Veld. 1997. Out-of-sample hedging effectiveness of currency futures for alternative models and hedging strategies. *Journal of Futures Markets* 17: 817–837.
- Dickey, D.A., and W.A. Fuller. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49: 1057–1072.
- Ederington, L.H. 1979. The hedging performance of the new futures markets. *Journal of Finance* 34: 157–170.
- Engle, R.F., and C.W. Granger. 1987. Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55: 251–276.
- Ghosh, A. 1993. Hedging with stock index futures: Estimation and forecasting with error correction model. *Journal of Futures Markets* 13: 743–752.
- Grammatikos, T., and A. Saunders. 1983. Stability and the hedging performance of foreign currency futures. *Journal of Futures Markets* 3: 295–305.
- Howard, C.T., and L.J. D'Antonio. 1984. A risk-return measure of hedging effectiveness. *Journal of Financial and Quantitative Analysis* 19: 101–112.
- Hsin, C.W., J. Kuo, and C.F. Lee. 1994. A new measure to compare the hedging effectiveness of foreign currency futures versus options. *Journal of Futures Markets* 14: 685–707.
- Hung, J.C., C.L. Chiu, and M.C. Lee. 2006. Hedging with zero-value at risk hedge ratio. *Applied Financial Economics* 16: 259–269.
- Johansen, S., and K. Juselius. 1990. Maximum likelihood estimation and inference on cointegration – With applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52: 169–210.
- Johnson, L.L. 1960. The theory of hedging and speculation in commodity futures. *Review of Economic Studies* 27: 139–151.
- Junkus, J.C., and C.F. Lee. 1985. Use of three index futures in hedging decisions. *Journal of Futures Markets* 5: 201–222.

- Kolb, R.W., and J. Okunev. 1992. An empirical evaluation of the extended mean-Gini coefficient for futures hedging. *Journal of Futures Markets* 12: 177–186.
- . 1993. Utility maximizing hedge ratios in the extended mean Gini framework. *Journal of Futures Markets* 13: 597–609.
- Kroner, K.F., and J. Sultan. 1993. Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis* 28: 535–551.
- Lee, C.F., J.E. Finnerty, J. Lee, A.C. Lee, and D.H. Wort. 2013. *Security analysis, portfolio management, and financial derivatives*. Hackensack: World Scientific Publishers.
- Lee, H.T., and J. Yoder. 2007. Optimal hedging with a regime-switching time-varying correlation GARCH model. *Journal of Futures Markets* 27: 495–516.
- Lence, S.H. 1995. The economic value of minimum-variance hedges. *American Journal of Agricultural Economics* 77: 353–364.
- . 1996. Relaxing the assumptions of minimum variance hedging. *Journal of Agricultural and Resource Economics* 21: 39–55.
- Lien, D. 1996. The effect of the cointegration relationship on futures hedging: A note. *Journal of Futures Markets* 16: 773–780.
- . 2004. Cointegration and the optimal hedge ratio: The general case. *The Quarterly Review of Economics and Finance* 44: 654–658.
- Lien, D., and X. Luo. 1993a. Estimating the extended mean-Gini coefficient for futures hedging. *Journal of Futures Markets* 13: 665–676.
- . 1993b. Estimating multiperiod hedge ratios in cointegrated markets. *Journal of Futures Markets* 13: 909–920.
- Lien, D., and D.R. Shaffer. 1999. Note on estimating the minimum extended Gini hedge ratio. *Journal of Futures Markets* 19: 101–113.
- Lien, D., and K. Shrestha. 2007. An empirical analysis of the relationship between hedge ratio and hedging horizon using wavelet analysis. *Journal of Futures Markets* 27: 127–150.
- . 2010. Estimating optimal hedge ratio: A multivariate skew-normal distribution. *Applied Financial Economics* 20: 627–636.
- Lien, D., and Y.K. Tse. 1998. Hedging time-varying downside risk. *Journal of Futures Markets* 18: 705–722.
- . 2000. Hedging downside risk with futures contracts. *Applied Financial Economics* 10: 163–170.
- Maghyreh, A.I., B. Awartani, and H. Abdoh. 2019. The co-movement between oil and clean energy stocks: A wavelet-based analysis of horizon associations. *Energy* 169: 895–913.
- Malliaris, A.G., and J.L. Urrutia. 1991. The impact of the lengths of estimation periods and hedging horizons on the effectiveness of a hedge: Evidence from foreign currency futures. *Journal of Futures Markets* 3: 271–289.
- Myers, R.J., and S.R. Thompson. 1989. Generalized optimal hedge ratio estimation. *American Journal of Agricultural Economics* 71: 858–868.
- Osterwald-Lenum, M. 1992. A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics* 54: 461–471.
- Phillips, P.C.B., and S. Ouliaris. 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica* 58: 165–193.
- Phillips, P.C.B., and P. Perron. 1988. Testing unit roots in time series regression. *Biometrika* 75: 335–346.
- Rutledge, D.J.S. 1972. Hedgers' demand for futures contracts: A theoretical framework with applications to the United States soybean complex. *Food Research Institute Studies* 11: 237–256.
- Sephton, P.S. 1993. Hedging wheat and canola at the Winnipeg commodity exchange. *Applied Financial Economics* 3: 67–72.
- Shalit, H. 1995. Mean-Gini hedging in futures markets. *Journal of Futures Markets* 15: 617–635.
- Sultan, J., A.K. Alexandridis, M. Hasan, and X. Guo. 2019. Hedging performance of multiscale hedge ratios. *Journal of Futures Markets* 39: 1613–1632.
- Working, H. 1953. Hedging reconsidered. *Journal of Farm Economics* 35: 544–561.



Credit Risk Modeling: A General Framework **75**

Ren-Raw Chen

Abstract

The two well-known approaches for credit risk modeling, structural and reduced form approaches, have their advantages and disadvantages. Due to the fundamentally different assumptions of the two approaches, the structural models are used for default prediction that focuses on equity prices and reduced form models are used for credit derivatives pricing that focuses on debt values. In this chapter, via a simple discrete binomial structure, we provide a unified view of the two approaches. In particular, in our formulation, the pricing formulas for risky debts are identical under the two approaches. The two approaches differ in only the recovery assumption. This result makes comparison of various models empirically possible. We demonstrate, in a credit derivative example that is sensitive to the recovery assumption, how different recovery assumptions impact its prices.

Keywords

Interest Rate · Credit Default Swap · Default Probability · Bond Price · Reduce Form Model

75.1 Introduction

There have been two well-known approaches, structural and reduced form, for credit risk modeling. Reduced form models, represented by Jarrow and Turnbull (1995) and Duffie and Singleton (1997, 1999) assume defaults (or credit events) occur

I would like to thank Jay Huang and Sanlin Chung for their helpful comments

R.-R. Chen (✉)

Fordham University, New York, NY, USA

e-mail: rchen@fordham.edu

exogenously (usually by a Poisson process) and a separately specified recovery is paid upon default.¹ Structural models, on the other hand, assume defaults occur when the value of the firm falls below a certain default point and a certain recovery is paid. The “true structural” model of Geske (1977) and Geske and Johnson (1984) assumes the default point to be the market value of debt that is endogenously computed and the firm value is the recovery. This formulation leads to multi-variate probability functions. The “barrier structural” models, pioneered by Black and Cox (1976), obtain uni-variate valuation formulas by assuming an exogenous default point and an exogenous recovery amount.²

Structural models for credit risk modeling have been mainly used for default prediction³ or capital structure analysis while reduced form models are mainly used by investment banks to price credit derivatives.⁴ This is because structural models rely on the information from equity prices while reduced form models from debt prices. Furthermore, reduced form models are more computationally efficient due to their exogenous default and recovery assumptions, which are important for pricing credit derivatives. Another reason for the reduced form models to be chosen by investment banks is its relative ease to incorporate the term structure of the default free interest rates.⁵

In this chapter, we clarify the difference between the “true structural” model of Geske (1977) and Geske and Johnson (1984) in which defaults occur when the (market) value of the firm falls below the (market) value of debt and the “barrier structural model” in which defaults occur when the value of the firm crosses an exogenously pre-defined barrier. The former valuation leads to multi-variate distributions while the latter is a univariate valuation.⁶ We then show that reduced-form and structural models can be made consistent under a simple discrete binomial formulation. Under this binomial formulation, the two sets of models have identical pricing formulas for risky debts. The only difference is different recovery assumptions. This result makes it possible to compare various models. In an application, we use credit default swaps to examine the impact of different recovery assumptions.

It is generally understood that structural models are difficult to calibrate. For example, for an n -period bond, we need n -dimensional probability functions, which

¹Other reduced form models include, among others, Duffie and Lando (1997), Jarrow et al. (1997), Jarrow and Yu (2001), Lando (1998), Madan and Unal (2000), and Schönbucher (1998).

²Other structural models include, for example, Anderson and Sundaresan (1996), Leland and Toft (1996), Longstaff and Schwartz (1995), Zhou (2001), and Bélanger et al. (2002). Readers can also find a more thorough survey by Uhrig-Homburg (2002).

³For example, see KMV's EDF and Moody's RiskCalc.

⁴It should be noted that structural models are also proposed for pricing risky bonds. However, they have never gained support from the industry due to the difficulty in calibration.

⁵The two most known reduced form models by Jarrow and Turnbull (1995) and Duffie and Singleton (1999) can both be easily incorporated into existing term structure models.

⁶Bélanger et al. (2002) provide a unified model that nests all barrier structural models. This chapter can be viewed as an extension of their work.

are computationally expensive. However, in this chapter, we employ the standard one-dimensional equity binomial model, which can be accurate to the second decimal place in 50 steps.

The remaining of the paper is organized as follows. Section 75.2 lays out the basic valuation equations. In particular, we specify the forward measure technique to incorporate stochastic interest rates. Section 75.3 presents the binomial framework under which all models share the same valuation formula. In Sect. 75.4, we discuss calibration issues of various models and perform model comparison. Section 75.5 concludes.

75.2 Basic Setup

Define a T -maturity default-free pure discount bond price at current time t , denoted $P(t, T)$ with $P(T, T) = 1$ for all T . Also define a set of dates: $0 = T_0 < T_1 < \dots < T_{N-1} < T_N$. For simplicity and without loss of generality, we also assume $h = T_i - T_{i-1}$ for all $i = 1, \dots, N$. Let $\ell(T_i, T_j)$, $0 \leq i < j \leq N$, represent the default free “term” interest rate (annualized) over $[T_i, T_j]$. Hence, by definition:

$$\ell(T_i, T_j) = \frac{1}{T_j - T_i} \left(\frac{1}{P(T_i, T_j)} - 1 \right) \quad (1)$$

Denote by Q the risk neutral measure under which defined the instantaneous short rate r . Then we have:

$$P(t, T) = E_t^Q[\Lambda(t, T)] \quad (2)$$

where

$$\Lambda(t, s) = \exp \left(- \int_t^s r(u) du \right), 0 \leq t < s \leq T \text{ and}$$

$E_t^Q[\cdot]$ represents the expectation conditional on the information set at time t under the risk neutral measure Q .

For the sake of convenience, we further define T_n -forward measure, denoted by F_n , to be equivalent to the measure Q and the corresponding Radon-Nikodym derivative is given by⁷:

$$\frac{dF_n}{dQ} = \frac{\Lambda(t, T_n)}{P(t, T_n)} \quad (3)$$

⁷For the forward measure, see, for example, Jamshidian (1987) and Hull (2000).

Under (3) the forward price is merely the forward expectation of the bond price:

$$F(0, t, s) = E_0^{F_t}[P(t, s)] = \frac{P(0, s)}{P(0, t)} \quad (4)$$

and so is the discrete forward rate:

$$E_0^{F_t}[l(u, t)h] = \frac{1}{F(0, u, t)} - 1 \quad (5)$$

Finally, we reiterate the separation of expectation using the forward measure:

$$\begin{aligned} E_0^Q[\Lambda(0, t)I_{\{t < \tau\}}] &= E_0^Q[\Lambda(0, t)]E_0^{F_t}[I_{\{t < \tau\}}] \\ &= P(0, t)Q(0, t) \end{aligned} \quad (6)$$

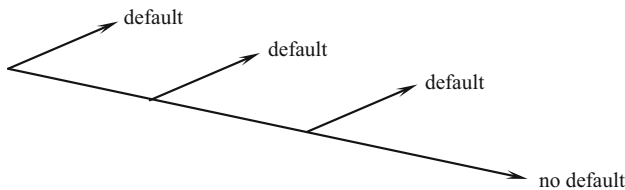
where τ is the default time and:

$$Q(0, t) = E_0^{F_t}[I_{\{t < \tau\}}] \quad (7)$$

as the survival probability under the forward measure where $I_{\{\cdot\}}$ is an indicator function and τ is the default time. Equation 6 represents the present of \$1 paid if there is no default and 0 otherwise. Hence, it is also known as the risky discount factor.

75.3 The Unified Model

The unified model is a simple binomial model that defines default and no default states at every period. This model is general to accommodate any form of recovery upon default and any form of cash flows under no default.



Hence, their solution to the risky fixed rate coupon bond can be written as:

$$\begin{aligned} V(0, T_n) &= E_0^Q \left[\sum_{i=1}^n \Lambda(0, T_i) c_n h I_{\{\tau > T_i\}} + \Lambda(0, T_n) I_{\{\tau > T_n\}} \right] \\ &\quad + \Lambda(0, \tau) w_n(A(\tau), r(\tau)) I_{\{\tau < T_n\}} \\ &= \sum_{i=1}^n P(0, T_i) Q(0, T_i) c_n h + P(0, T_n) Q(0, T_n) + R_n(0) \end{aligned} \quad (8)$$

where c_n is the fixed coupon, $w_n(\cdot, \cdot)$, which is a function of some state variable A and interest rate r , represents the recovery amount upon default, and $R_n(0)$ is the current value of expected recovery. Note that the first two terms of the second line follows directly from (6) and (7). They represent the value of coupons and face value. The last term represents the current recover value, which is an expected present value of the recovery amount upon default, a function of some state variable(s), A and the risk free interest rate, r .

The solution to the risky floating rate bond is:

$$\begin{aligned}
 V^*(0, T_n) &= E_0^Q \left[\sum_{i=1}^n \Lambda(0, T_i) \ell(T_{i-1}, T_i) h I_{\{\tau > T_i\}} + \Lambda(0, T_n) I_{\{\tau > T_n\}} \right. \\
 &\quad \left. + \Lambda(0, \tau) w_n(A(\tau), r(\tau)) I_{\{\tau < T_n\}} \right] \\
 &= \sum_{i=1}^n P(0, T_i) Q(0, T_i) \left[F(0, T_{i-1}, T_i)^{-1} - 1 \right] h \gamma_{i-1, i} \\
 &\quad + P(0, T_n) Q(0, T_n) + R_n(0)
 \end{aligned} \tag{9}$$

where

$$\gamma_{i-1, i} = \text{cov}[\ell(T_{i-1}, T_i), I_{\{\tau > T_i\}}]$$

is the covariance under the forward measure between the random interest rate and the default event, which is usually assumed 0. It is generally accepted that default probability is positively correlated with interest rates; but default events (or equivalently, default time) are independent of interest rates. The second line of (9) comes directly from (5). As we can see, the difference between a floater and a fixed rate bond is the coupon payment series, where the former is taken from the forward curve and the latter is a fixed constant. Hence, for the rest of the paper, we simply concentrate on fixed rate bond.

It is not unreasonable to assume that defaults can occur only at discrete points (i.e. coupon payment times). Usually the company does not have to declare default if there is no cash pressure. This assumption validates the use of the Geske-Johnson structural model. So far, we have not used any model specification. The remainder of this section, presents $V(0, T_n)$ in various model specifications.

75.3.1 The Jarrow-Turnbull Model

The Jarrow-Turnbull model assumes that a fixed recovery is paid at maturity regardless of the time of default. Hence, the closed form solution exists for the coupon bond as follows:

$$V_{JT}(0, T_n) = \sum_{i=1}^n P(0, T_i) Q(0, T_i) c_n h + P(0, T_n) Q(0, T_n) + w_n P(0, T_n) \times [1 - Q(0, T_n)] \quad (10)$$

In such a case, $R_n(0) = w_n P(0, T_n) [1 - Q(0, T_n)]$. This model is also recognized as “recovery of proportional to par.” An extended Jarrow-Turnbull model of the following is usually used in the industry in which defaults are allowed only at coupon times and a fixed recovery is paid upon default.

$$\Phi_{JT}(0, T_n) = \sum_{i=1}^n P(0, T_i) Q(0, T_i) c_n h + P(0, T_n) Q(0, T_n) + w_n(T_i) P(0, T_i) [Q(0, T_{i-1}) - Q(0, T_i)] \quad (11)$$

In such a case, $R_n(0) = \sum_{i=1}^n w_n(T_i) P(0, T_i) [Q(0, T_{i-1}) - Q(0, T_i)]$. In many cases, continuous default is necessary, hence a continuous version of recovery is also commonly used: $R_n(0) = \int_0^{T_n} w_n(t) P(0, t) [-dQ(0, t)]$.

75.3.2 The Duffie-Singleton Model

The Duffie-Singleton model assumes recovery is paid immediately upon default and equals a fraction of what the bond is worth immediately prior to default. In our formulation, it means:

$$w_n(t) = \delta V(t, T_n) \quad (12)$$

where δ is the constant recovery ratio on the value of the bond prior to default. Substituting this result back to (8), we get:

$$V_{DS}(0, T_n) = \sum_{i=1}^n P(0, T_i) Q^*(0, T_i) c_n h + P(0, T_n) Q^*(0, T_n) \quad (13)$$

where c_n is the fixed coupon and

$$Q^*(0, T_i) = E_0^{F_i} \left[e^{-(1-\delta) \int_0^{T_i} \lambda(t) dt} \right] = \begin{cases} Q(0, T_i) e^{\delta \int_0^{T_i} \lambda(t) dt} \\ Q(0, T_i)^{1-\delta} \end{cases} \quad (14)$$

where $\lambda(t)$ is the Poisson intensity parameter.⁸ Default occurs when there is a jump with a probability $\lambda(t)dt$ over the period dt . It can be seen that the Duffie-Singleton model cannot differentiate the survival probability from the recovery, a drawback of the model. However, the solution represented by (13) and (14) is a closed form solution of the affine style, more easily to derive closed form solutions when the recovery ratio and the intensity parameter are random.

⁸If the intensity parameter and/or the recovery rate are random, then (14) becomes:

It is clear that in the Duffie-Singleton model, recovery is blended into survival probabilities. In other words, recovery in the Duffie-Singleton model contains survival probabilities. Formally, we can write recovery as:

$$R_n(0) = \sum_{i=1}^n P(0, T_i) Q(0, T_i) [M(0, T_i) - 1] c_n h + P(0, T_n) Q^*(0, T_n) \quad (15)$$

where⁹

$$M(0, t) = e^{\delta \int_0^t \lambda(u) du}$$

75.3.3 The Extended Merton Model

The structural models can be traced back to Black and Scholes (1973) and Merton (1974) who observe that the company's equity is a European call option and hence the single-maturity-date debt contains default risk identical to a covered call. The recovery in the Black-Scholes-Merton model is therefore the firm value if default occurs at maturity. To extend the single period model of Black-Scholes and Merton, as mentioned earlier, there are two approaches. The “barrier structural” models assume an exogenous default barrier. Default is defined as the asset value crossing such a barrier.¹⁰ The extension along this is pioneered by Black and Cox (1976), followed by Leland and Toft (1996) and Longstaff and Schwartz (1995), and recently extended by Bélanger et al. (2002). Another is a “pure structural” model by Geske (1977) and Geske and Johnson (1984) who adopt the compound option approach and treat default as the inability of the company to fulfill its debt obligations (i.e. negative equity value).

In this sub-section, we include multi-period debt structure in the Merton model. We assume defaults occur if the firm fails to meet its coupon obligations at any give time. Since the coupon obligations are exogenously given, the “barrier structural” models can be viewed as the continuous time limit of what is described here.

Consistent with our discrete setup, let $X_1, \dots, X_n$ be the series of external barriers, crossing which by the firm value represents default. These discrete time barriers can be interpreted as cash obligations at each time and failure to make the obligation results in default of the firm. These cash obligations can be regarded as a series of zero coupon bonds issued by the firm. The values of assets and debts (zero coupon) are labeled as $A(t)$, and $D(t, T_n)$ respectively for an arbitrary t . Default is defined as

⁹If the intensity parameter and/or the recovery rate are random, then we have:
where

¹⁰The barrier-based structural models are particularly popular in industry. See, for example, KMV (recently acquired by Moody's) and CreditGrades.

$A(T_i) < X_i$ at time T_i . To obtain closed form solutions, we assume the continuity of $A(t)$, as by Geske (1977) and Geske and Johnson (1984). To arrive at closed form results, we need to assume a log normal process for the firm's asset value and a normal process for the instantaneous short rate:

$$\begin{aligned} \begin{bmatrix} \frac{dA(t)}{A(t)} \\ dr(t) \end{bmatrix} &= \begin{bmatrix} r(t) \\ \mu(r, t) \end{bmatrix} dt + \begin{bmatrix} \sigma_A & 0 \\ 0 & \sigma_r \end{bmatrix} \\ &\quad \times \begin{bmatrix} \sqrt{1 - \rho_{Ar}^2} & \rho_{Ar} \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} dW_A^Q(t) \\ dW_r^Q(t) \end{bmatrix} \end{aligned} \quad (16)$$

where σ_A , σ_r , and ρ_{Ar} are constants, and $dW_A^Q dW_r^Q = 0$. From (16), we know that the correlation between the (log) asset value and the interest rate is ρ_{Ar} . Also note that both $W_A^Q(t)$ and $W_r^Q(t)$ are Wiener processes under the Q measure. Applying Ito's lemma to (2) in Sect. 75.2, we obtain that:

$$\begin{aligned} \frac{dP(t, T)}{P(t, T)} &= \frac{1}{P(t, T)} \left[\frac{\partial P(t, T)}{\partial r} \mu(r, t) + \frac{1}{2} \frac{\partial^2 P(t, T)}{\partial r^2} \sigma_r^2 + \frac{\partial P(t, T)}{\partial t} \right] dt \\ &\quad + \frac{1}{P(t, T)} \frac{\partial P(t, T)}{\partial r} \sigma_r dW_r^Q(t) \\ &\equiv r(t)dt + \sigma_P(r, t, T) dW_r^Q(t) \end{aligned} \quad (17)$$

The Gaussian models that satisfy (17) are Vasicek (1977), Ho and Lee (1986), and Hull and White (1990), in all of which the diffusion term, $\sigma_P(r, t, T)$, is independent of the interest rate and can be written as $\sigma_P(t, T)$.

Then, we can derive the value of T_n -maturity debt is:

$$\begin{aligned} D(0, T_n) &= E_0^Q[\Lambda(0, T_n) \min \{X_n, A(T_n)\}] \\ &= P(0, T_n) E_0^{F_n}[\min \{X_n, A(T_n)\}] n > 1 \\ &= P(0, T_n) X_n N_n^- + A(0) (N_{n-1}^+ - N_n^+) \end{aligned} \quad (18)$$

where

$$\begin{aligned} N_j^\pm &= N_j \left(h_1^\pm(X_1), \dots, h_j^\pm(X_j); \rho_{ij} \right) \\ \rho_{ij} &= \sqrt{T_i/T_j}, i < j \\ h_i^\pm(X_i) &= \frac{\ln \frac{A(0)}{P(0, T_i) X_i} \pm \frac{1}{2} v^2(0, T_i)}{v(0, T_i)} \\ v^2(0, T_i) &= \int_0^{T_i} \sigma_A^2 + \sigma_P^2(u, T_i) - 2\rho_{Ar} \sigma_A \sigma_P(u, T_i) du \end{aligned}$$

where ρ_{ij} represents the auto-correlation between $\ln A(T_i)$ and $\ln A(T_j)$ which is $\sqrt{T_i/T_j}$ for $T_i < T_j$. The derivation of (18) is given in an appendix. In (18), $N_j(\cdot \cdot)$ is a j -dimensional cumulative normal probability.¹¹ The derivation of (18) is given in an appendix that both N_n^+ and N_n^- are probabilities being in the money, only defined in different probability measures.

Equation 18 has an interesting interpretation. Note that N_n^- is the survival probability from now till T_n and $N_{n-1}^- - N_n^-$ (or $N_{n-1}^+ - N_n^+$ under a different measure) is the *unconditional* default probability between T_{n-1} and T_n . In other words, (18) implies that if default does not happen (with probability N_n^-), the bond receives X_n . If default does happen, it receives $A(T_n)$. Note that any prior default (default before T_{n-1}) should pay no recovery for such a bond since it is the most junior bond and the last to receive recovery. And this recovery is multiplied by the default probability and discounted to be $A(0)(N_{n-1}^+ - N_n^+)$ today.

It should be noted that (18) is not a closed form result unless $v(0, T_i)$ has a closed form expression. Rabinovitch (1989) shows that a closed form expression for $v(0, T_i)$ exists if the term structure model follows Vasicek (1977) and does not exist if it follows Cox-Ingersoll-Ross (1985). We restate the closed form result of Rabinovitch in an appendix.

The T_n -maturity coupon bond that pays hc_n as coupons can be regarded as the total debt of the firm if $X_i = hc_n$ for $i = 1, \dots, n-1$ and $X_n = 1 + hc_n$ where h is defined in (18), as the time period length:

$$\begin{aligned} V_{EM}(0, T_n) &= \sum_{i=1}^n D(0, T_i) \\ &= A(0)[1 - N_n^+] + \sum_{i=1}^n P(0, T_i)X_iN_i^- \\ &= R_n(0) + \sum_{i=1}^n P(0, T_i)Q(0, T_i)hc_n \end{aligned} \quad (19)$$

where the forward survival probability notation $Q(t, T_i)$ replaces multi-variate normal probabilities N_i^- . Eq. 19 resembles (8) remarkably. The only difference is the recovery assumption. Note that the first term $1 - N_n^+$ represents the total unconditional default probability (under a different measure). While (19) is much simpler than (8), it requires the knowledge the underlying asset. But this problem can be easily resolved by a recursive equation:

$$\begin{aligned} &\frac{V_{EM}(0, T_n) - \sum_{i=1}^n P(0, T_i)Q(0, T_i)c_n - P(0, T_n)Q(0, T_n)}{1 - N_n^+} \\ &= \frac{V_{EM}(0, T_{n-1}) - \sum_{i=1}^{n-1} P(0, T_i)Q(0, T_i)c_{n-1} - P(0, T_{n-1})Q(0, T_{n-1})}{1 - N_{n-1}^+} \end{aligned} \quad (20)$$

¹¹ Equations 17 and 18 are consistent with Vasicek (1977), but not Cox et al. (1985).

Under the par bond assumption, $V(0, T_n) = 1$ and $y_n = c_n$ and hence (20) can be written as:

$$\begin{aligned} & \frac{1 - \sum_{i=1}^n P(0, T_i) Q(0, T_i) h y_n - P(0, T_n) Q(0, T_n)}{1 - \sum_{i=1}^{n-1} P(0, T_i) Q(0, T_i) h y_{n-1} - P(0, T_{n-1}) Q(0, T_{n-1})} \\ &= \frac{1 - N_n^+}{1 - N_{n-1}^+} = \theta_n \end{aligned} \quad (21)$$

Note the following identity:

$$1 = \sum_{i=1}^n P(0, T_i) Q(0, T_i) h c_n^0 + P(0, T_n) Q(0, T_n) \quad (22)$$

where $c_n^0 = y_n^0$ is the coupon of a zero-recovery risky par bond. Hence, (17) can be simplified as follows:

$$\frac{\sum_{i=1}^n P(0, T_i) Q(0, T_i) (y_n^0 - y_n)}{\sum_{i=1}^{n-1} P(0, T_i) Q(0, T_i) (y_{n-1}^0 - y_{n-1})} = \theta_n \quad (23)$$

The extended Merton model is usually categorized as a structural model, because of its endogenous recovery assumption and the use of asset value. However, default is defined as the asset value crossing an external barrier. This could cause negative equity value. To see that, we can look at the payoffs at time T_{n-1} :

$$\begin{aligned} & A(T_{n-1}) > X_{n-1} \\ \Rightarrow & \begin{cases} D(T_{n-1}, T_{n-1}) = X_{n-1} \\ D(T_{n-1}, T_n) = P(T_{n-1}, T_n) E_{n-1}^E [\min \{A(T_n), X_n\}] \\ E(T_{n-1}) = P(T_{n-1}, T_n) E_{n-1}^F [\max \{A(T_n) - X_n, 0\}] - X_{n-1} \end{cases} \\ & A(T_{n-1}) \leq X_{n-1} \\ \Rightarrow & \begin{cases} D(T_{n-1}, T_{n-1}) = A(T_{n-1}) \\ D(T_{n-1}, T_n) = 0 \\ E(T_{n-1}) = 0 \end{cases} \end{aligned}$$

When $A(T_{n-1})$ is small, there is no guarantee that the equity value, $E(T_{n-1})$, can exceed X_{n-1} since X_{n-1} can be arbitrary. To keep the continuity assumption of the asset value at time T_{n-1} , we need to issue new equity when it is negative. In other words, we allow the company to raise new equity when it is already in bankruptcy. Clearly this is not possible in reality. There are three approaches to avoid such a problem. The first is to set the default boundary at equity value, not asset value, equal to X_{n-1} . This is the Geske-Johnson model that we will discuss in the next sub-section. Second, we should treat the underlying asset as an unobservable state variable and specify recover value separately (e.g. Longstaff and Schwartz (1995) and Zhou (2001)). But doing so effectively transforms the structural model into

reduced form in that both defaults and recovery are exogenously specified. It differs from the reduce form approach only by different default process, one assumes a Poisson process and the other assumes a diffusion variable crossing a barrier. Third, we can simplify the debt structure so that an endogenous barrier can be solved (e.g. Leland 1994; Leland and Toft 1996).

75.3.4 The Geske-Johnson Model

A true structural model is Geske (1977) and Geske and Johnson (1984) in which internal strikes are solved to guarantee positive equity value. We extend the Geske-Johnson model of an n period risky debt to incorporate random interest rates.^{12,13} The T_n -maturity zero coupon bond can be written as follows.

$$D(0, T_n) = \sum_{i=1}^n P(0, T_i) X_i [\Pi_i^-(K_{1n}, \dots, K_{in}) - \Pi_i^-(K_{1n-1}, \dots, K_{in-1})] + A(0) [\Pi_{n-1}^+(K_{1n-1}, \dots, K_{n-1n-1}) - \Pi_n^+(K_{1n}, \dots, K_{nn})] \quad (24)$$

where¹⁴

$$\begin{aligned} \Pi_i^\pm(K_{1j}, K_{2j}, \dots, K_{ij}) = & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} N_i(h_1^\pm(K_{1j}(r)), h_2^\pm(K_{2j}(r)) \\ & \dots, h_i^\pm(K_{ij}(r))) \varphi(r(T_1), \\ & r(T_2), \dots, r(T_i)) \\ & dr(T_1) dr(T_2), \dots, dr(T_i) \end{aligned}$$

for $n \geq j \geq i$ where φ is the joint density function of various interest rate levels observed at different times under the forward measure. Note that $h_j^\pm(K_{ij})$, for $i < j$, is to plug into (18) K_{ij} for the strike and K_{ij} is the internal solution to:

$$A(T_i) = \sum_{k=i}^j D(T_i, T_k) + X_i$$

¹²The formulas provided by Geske (1977) are incorrect and corrected by Geske and Johnson (1984). However, Geske and Johnson only present formulas for $n = 2$. Here, we generalize their formulas to an arbitrary n .

¹³Later on, for calibration, we also augment the model to include non-constant volatility.

¹⁴Note that the implementation of (24) does not need multi-variate integrals. The easiest way to implement it is to construct a bi-variate lattice. Eom et al. (2002), for example, use a one-dimension binomial model to implement the deterministic Geske-Johnson model.

which is a function of the interest rate at time T_i , and $K_{ji} = X_i$. All strikes are solved internally. Note that since interest rates are random, the internally solved strike, K_{ij} , is a function of r under the forward measure. If interest rates are deterministic, then $\Pi_i^\pm = N_i(\cdot \cdot \cdot)$ the standard multi-variate normal probability function. If we further assume that default points are equal to cash obligations, i.e. $K_{ij} = X_i$ for all j , then $\Pi_i^\pm = N_i^\pm$ defined in (18).

Although each bond is computed by a complicated formula, the coupon bond is not:

$$\begin{aligned} V_{GJ}(0, T_n) &= \sum_{i=1}^n D(0, T_i) \\ &= A(0) [1 - \Pi_n^+(K_{1n}, K_{2n}, \dots, K_{nn})] \\ &\quad + \sum_{i=1}^n P(0, T_i) X_i \Pi_i^-(K_{1n}, \dots, K_{in}) \\ &= R_n(0) + \sum_{i=1}^n P(0, T_i) Q(0, T_i) hc_n \\ &\quad + P(0, T_n) Q(0, T_n) \end{aligned} \quad (25)$$

where $X_i = hc_n$ for $i < n$ and $X_i = 1 + c_n h$ for $i = n$. This equation is extremely similar to (19) except that the probabilities are defined differently due to different strikes. Hence, by observation, the recovery must be $R_n(0) = A(0) \times [1 - \Pi_n^+(K_{1n}, K_{2n}, \dots, K_{nn})]$. It is interesting to see that the Geske-Johnson model satisfies the condition $Q(0, T_i) > Q(0, T_{i+1})$.

Note again that $\Pi_n^-(K_{1n}, K_{2n}, \dots, K_{nn})$ is the total survival probability (under the forward measure) because the asset value, $A(t)$ needs to stay above its default boundaries, K_{ij} at all times. Hence, the total (cumulative) default probability is $[1 - \Pi_n^-(K_{1n}, K_{2n}, \dots, K_{nn})]$. To see it, we know that the total default probability can be derived directly from summing the default probability of each period:

$$\begin{aligned} &\int_0^{K_{1n}} \varphi(A_1) dA_1 + \int_{K_{1n}}^\infty \int_0^{K_{2n}} \varphi(A_1, A_2) dA_2 dA_1 + \dots \\ &+ \int_{K_{1n}}^\infty \dots \int_{K_{n-1n}}^\infty \int_0^{K_{nn}} \varphi(A_1, \dots, A_n) dA_n \dots dA_1 \\ &= 1 - \int_{K_{1n}}^\infty \dots \int_{K_{n-1n}}^\infty \int_{K_{nn}}^\infty \varphi(A_1, \dots, A_n) dA_n \dots dA_1 \\ &= 1 - \Pi_n^-(K_{1n}, \dots, K_{nn}) \end{aligned} \quad (26)$$

where $A_i = A(T_i)$ is a short-hand notation. This is analogous to $\int_0^T -dQ(t) = 1 - Q(T)$. The recovery is to consider the cash amount received upon default:

$$\begin{aligned}
& P(0, T_1) \int_0^{K_{1n}} A_1 \varphi(A_1) dA_1 + P(0, T_2) \int_{K_{1n}}^\infty \int_0^{K_{2n}} A_2 \varphi(A_1, A_2) dA_2 dA_1 \\
& + \cdots + P(0, T_n) \int_{K_{1n}}^\infty \cdots \int_{K_{n-1n}}^\infty \int_0^{K_{nn}} A_n \varphi(A_1, \dots, A_n) dA_n \cdots dA_1 \\
& = A_0 \left[1 - \int_{K_{1n}}^\infty \cdots \int_{K_{n-1n}}^\infty \int_{K_{nn}}^\infty \varphi^+(A_1, \dots, A_n) dA_n \cdots dA_1 \right] \\
& = A_0 [1 - \Pi_n^+(K_{1n}, \dots, K_{nn})]
\end{aligned} \tag{27}$$

The change of measure can be found in an appendix.

In this unified framework, we can see that the difference in the Geske-Johnson model differs from Jarrow-Turnbull model is that Jarrow and Turnbull assume a fixed recovery value at maturity while Geske and Johnson assume a fixed recovery value at current time.

75.4 Calibration and Model Comparison

In this section, we demonstrate how different models can calibrate to the same market data and imply different parameter values. Equation 8 represents a general formula for all risky bond pricing models, reduced form and structural models included. It is seen that the difference only lies in the recovery assumption: recovery of the face value yields the Jarrow-Turnbull model; recovery of the market value yields the Duffie-Singleton model, and recovery of the asset value yields the Geske-Johnson model.

We shall first use a two-period model as an example to demonstrate the computational details. Then we provide analysis in a multi-period setting. The Vasicek model (1977) for the risk free term structure of interest rates is considered.

75.4.1 A Two-Period Example

We need three pieces of information to complete the calibration: risk free zero yield curve, a set of risky bond prices, and a recovery assumption. In the case of structural models, the recovery assumption is replaced by the volatility curve for the set of bond prices.

To demonstrate the calibration procedures for both reduced form and structural models, we use a two-period example. The following table describes the base case:

Time	1	2
Coupon	10	10
Face	100	100
Bond price	100	100
Yield curve	5%	5%

The Jarrow-Turnbull model can be calibrated as follows. From (11), we obtain the 1 year bond formula:

$$V_{JT}(0, 1) = P(0, 1)[Q(0, 1)(1 + c_1) + (1 - Q(0, 1))w_1]$$

We assume a fixed recovery rate of 0.4 of the principal and accrued interest. In this 1-year example, the recovery amount is \$44 if default occurs. Given coupon $c_1 = \$10/\$100 = 10\%$ and 1-year discount $P(0, 1) = e^{-5\%} = 0.9512$, we can solve for the survival probability to be $Q(0, 1) = 0.9262$. Again, from (11), for the 2-year bond:

$$V_{JT}(0, 2) = P(0, 1)Q(0, 1)c_2 + P(0, 2)Q(0, 2)(1 + c_2) \\ + P(0, 1)[1 - Q(0, 1)]w_2 + P(0, 2)[Q(0, 1) - Q(0, 2)]w_2$$

With the knowledge of $Q(0, 1)$, together with $P(0, 2) = e^{-10\%} = 0.9048$, we can then solve for the second period survival probability to be $Q(0, 2) = 0.8578$.

The Duffie-Singleton model can be calibrated as follows. From (13), we have:

$$V_{DS}(0, 1) = P(0, 1)Q^*(0, 1)(1 + c_1)$$

for the 1-year bond and

$$V_{DS}(0, 2) = P(0, 1)Q^*(0, 1)c_2 + P(0, 2)Q^*(0, 2)(1 + c_2)$$

for the 2-year bond. Hence, we solve for $Q^*(0, 1) = 0.9557$ and $Q^*(0, 2) = 0.9134$. From (14), we know that $Q^*(0, t) > Q(0, t)$ always due to non-negative recovery. By assuming a recovery rate of 0.4, we get $Q(0, 1) = 0.8929$ and $Q(0, 2) = 0.7973$, which are both lower than the survival probabilities calculated by the Jarrow-Turnbull model. Note that in both models, the recovery present values for 1 and 2-year bonds are 3.09 and 5.81 for Jarrow-Turnbull and 6.57 and 12.15 for Duffie-Singleton, respectively. It is seen that the recovery amounts for Duffie-Singleton are higher, and hence to maintain the same the bond values, the survival (default) probabilities of Duffie-Singleton have to be lower (higher) to balance it out.

In both the Jarrow-Turnbull model and the Duffie-Singleton model, given that there is no other random factor other than the interest rate, there is no need to identify a specific term structure model, given that survival and default probabilities are computed under the forward measure. However, in the Geske-Johnson model, in addition to random interest rates, there is a random “asset price” state variable. As a result, a specific term structure model needs to be specified in order to carry out survival and default probabilities.

To simplify the calculation and without any loss of generality, we assume a deterministic yield curve in this section. The Vasicek term structure model is

assumed in the next sub-section. In the Geske-Johnson model, the recovery amount is random and endogenous. Under deterministic interest rates, the one-period Geske-Johnson model is a Black-Scholes-Merton solution:

$$V_{GJ}(0, 1) = D(0, 1) = P(0, 1)X_1N(h_1^-(X_1)) + A(0)[1 - N(h_1^+(X_1))] \quad (28)$$

where $h_1^\pm$ is defined in (18). Since interest rates are non-stochastic, $P(0, 1) = \Lambda(0, 1)$. Equation 28 is identical to (18) when $\sigma_P = 0$ and $T_n = T_1 = 1$. For the 2-year bond, i.e. $T_n = T_2 = 2$ and $T_1 = 1$, we have two zero bond components:

$$D(0, 1) = P(0, 1)X_1N(h_1^-(X_1)) + A(0)[1 - N(h_1^+(X_1))] \quad (29)$$

and

$$\begin{aligned} D(0, 2) &= P(0, 1)X_1[N_1(h_1^-(K_{12})) - N_1(h_1^-(X_1))] \\ &\quad + P(0, 2)X_2N_2(h_1^-(K_{12}), h_2^-(X_2); \sqrt{1/2}) \\ &\quad + A(0)[N_1(h_1^+(X_1)) - N_2(h_1^+(K_{12}), h_2^+(X_2); \sqrt{1/2})] \end{aligned} \quad (30)$$

where $h_i^\pm$ for $i = 1, 2$ is defined in (18), $X_1 = c_2$, $X_2 = 1 + c_2$. Recall that K_{12} is the solution to $A(1) = D(1, 2) + X_1$ where $D(1, 2)$ is a bond price at $t = 1$. Finally, the 2-year coupon bond value is the sum of (29) and (30):

$$\begin{aligned} V_{GJ}(0, 2) &= D(0, 1) + D(0, 2) \\ &= P(0, 1)X_1N(h_1^-(X_1)) + P(0, 2)X_2N_2(h_1^-(K_{12}), h_2^-(X_2)) \\ &\quad + A_0[1 - N_2(h_1^+(K_{12}), h_2^+(X_2); \sqrt{1/2})] \end{aligned} \quad (31)$$

In the equation, we need to evaluate three probabilities. The first one is $N(h_1^-)$ which has already been evaluated. The second one is a bivariate normal probability with two separate strikes, K_{12} that has to be internally solved and X_2 that is the face and coupon value at maturity.

Note that X_1 is the first cash flow of the company, hence $X_1 = 10 + 110 = 120$. The value of $D(0, 1)$ hence contains the cash flow of the first bond and the coupon amount of the second bond. Here, we assume that the split of $D(0, 1)$ is proportional. That is the value of the 1-year bond is $\frac{110}{120}D(0, 1)$, and the value of the 2-year bond is $\frac{10}{120}D(0, 1) + D(0, 2)$.

In the original Geske-Johnson model, the volatility is assumed flat (as in Black-Scholes 1973). Unfortunately under this condition, the calibration of the second bond becomes impossible. Hence, we extend the model to include a volatility curve, i.e. $v(0, 2)^2 = v(0, 1)^2 + v(1, 2)^2$. This flexibility allows us to calibrate the model to the 2-year bond price. The results are $A(0) = 2351.65$, $v(0, 1) = 1.5$, and $v(1, 2) = 0.69$. Under such results, the survival probabilities are $Q(0, 1) = N_1^-(K_{12}) = 0.9426$ and $Q(0, 2) = N_2^-(K_{12}, X_2) = 0.8592$ respectively. And the recovery values for the two bonds are computed by subtracting the coupon value

from the corresponding discount debt value. For example, $D(0, 1) = 109.09$ which is split into two parts – the 1-year bond of \$100 and the coupon of the 2-year bond of \$9.09. The 1-year bond has a coupon portion of $110 \times Q(0, 1) \times P(0, 1) = 110 \times 0.9426 \times 0.9512 = 98.63$, and hence has a recovery value of $100 - 98.63 = 1.37$. A similar calculation gives the recovery value of the 2-year bond as 5.52.¹⁵

For easy comparison, we put together all the numbers in the following table. We observe that the Duffie-Singleton (DS) model has the lowest survival probabilities and therefore should have the highest recovery values, due to the fact that they both contribute positively to the bond price. The Geske-Johnson (GJ) model has the highest survival probabilities and lowest recovery values. The Jarrow-Turnbull (JT) model is in between. Since the bond price is traded at par, higher survival probabilities need to be balanced by lower recovery values.

	JT	DS	GJ
$Q(0,1)$	0.9262	0.8929	0.9426
$Q(0,2)$	0.8578	0.7923	0.8592
Total recovery value (i.e. $R(0)$)	8.90	18.68	6.89
Recovery of first bond	3.09	6.53	1.37
Recovery of second bond	5.81	12.15	5.52

75.4.2 Multi-period Analysis

In this section, we examine the multi-period behavior of reduced form models, namely Jarrow-Turnbull and Duffie-Singleton and the structural model of Geske-Johnson. We examine the default and recovery implication under an n -period setting. We also incorporate stochastic interest rates in our analysis. We replace the interest rate process in (16) by the following Ornstein-Uhlenbeck process:

$$dr(t) = \alpha \left(\mu - \frac{\delta q}{\alpha} - r(t) \right) dt + \delta dW_r^Q(t) \quad (32)$$

where the parameter values are given for flat (constant) and upward sloping yield curves as follows.

Vasicek model parameters		
	Flat	Upward
α (reverting speed)	∞	0.40
μ (reverting level)	0.05	0.08
δ (volatility)	0.00	0.05
q (mkt. price of risk)	0.00	-0.10
$r(0)$ (initial rate)	0.05	0.05

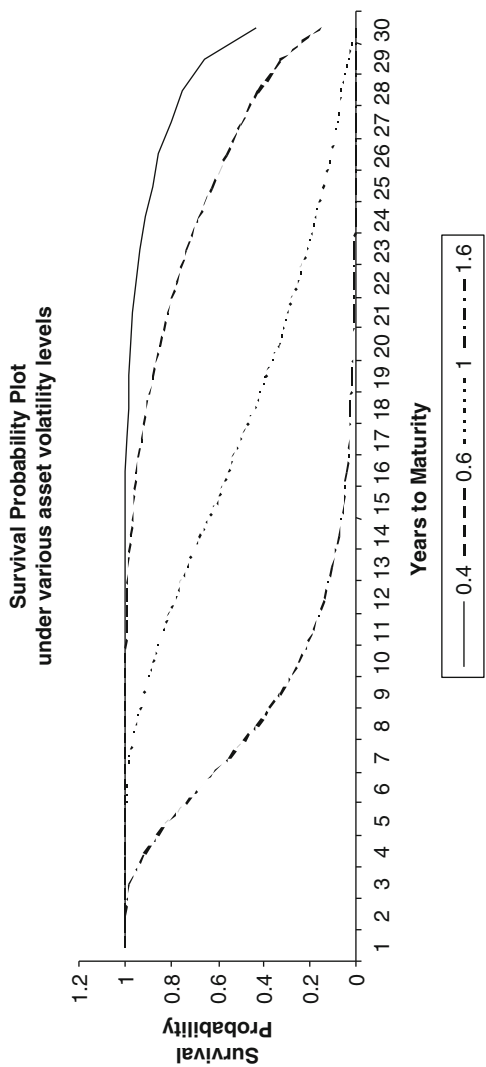
¹⁵ Note that the total recovery value needs to be 6.89 by (31).

We first examine flat yield curve. Then we use the upward sloping parameter values to construct the Vasicek model. The two sets of results are similar and we shall present the flat yield curve result and the other one is available upon request.

We first examine the case of extremely low coupons. This is the case where we can see the fundamental difference between the structural model of Geske-Johnson and the reduced form models of Jarrow-Turnbull and Duffie-Singleton. Note that the Jarrow-Turnbull model is (11) and the Duffie-Singleton model is (13) and (14), both with $c_n = 0$. The Geske-Johnson model is (25) with $X_i = 0$ for $i < n$. In all the models, the zero coupon bond price, $P(t, T)$, should follow the Vasicek model (formula given in an appendix.) The face value of debt is 110. We run the Geske-Johnson model with an asset value of $A(0) = 184$ and various volatility levels: 0.4, 0.6, 1.0, and 1.6. The results computed are summarized as follows.

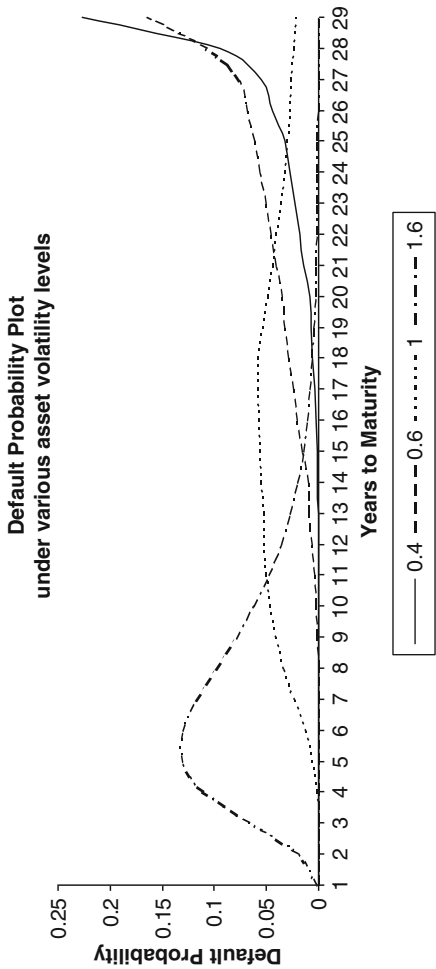
GJ Model				
Volatility	0.4	0.6	1.0	1.6
Equity value	169.25	177.92	183.49	183.94
Debt value	14.74	6.08	0.51	0.06
Recovery	4.06	2.08	0.19	0.00
JT Model				
Recovery rate	3.69%	1.89%	0.17%	0.00%
Intensity	1.79%	4.85%	13.19%	20.00%

As the volatility goes up, the equity value in the Geske-Johnson model goes up (i.e. call option value goes up.) Since the asset value is fixed at 184, the debt value goes down. The survival probability curves under various volatility scenarios of the Geske-Johnson model are plotted in Fig. 1; and the default probability curves (unconditional, i.e. $Q(T_{i-1}) - Q(T_i)$) are plotted in Fig. 2. We observe several results. First, as the risk of default becomes eminent (i.e. high volatility and low debt value), the likelihood of default shifts from far term (peak at year 30 for volatility = 0.4) to near term (peak at year 5 for volatility = 1.6). Second, it is seen that the asset volatility has a huge impact on the shape of the survival probability curve. The Geske-Johnson model is able to generate humped default probability curve, often observed empirically. Third, these differently shaped probability curves are generated by one single debt, something not possible reduced form models. Both the Jarrow-Turnbull and Duffie-Single models cannot generate such probability curves with one single bond, due to the lack of information of intermediate cash flows. Corresponding to the recovery amounts under the Geske-Johnson model, we set the fixed recovery rate of the Jarrow-Turnbull model as shown in the above table. Given that one bond can only imply one intensity parameter value, we set it (under each scenario) so that the zero coupon bond price generated by the Geske-Johnson model is matched. As shown in the above table, the intensity rate goes from 1.79% per annum to 20% per annum. Note that flat intensity value is equivalent to a flat *conditional* default probability curve. To visualize the difference this flat conditional default probability curve of the Jarrow-Turnbull model with the non-flat curve generated by the Geske-Johnson model, we plot in Fig. 3 the case of volatility = 1.6. The conditional default probabilities are calculated as $\frac{Q(T_{i-1}) - Q(T_i)}{Q(T_{i-1})}$.



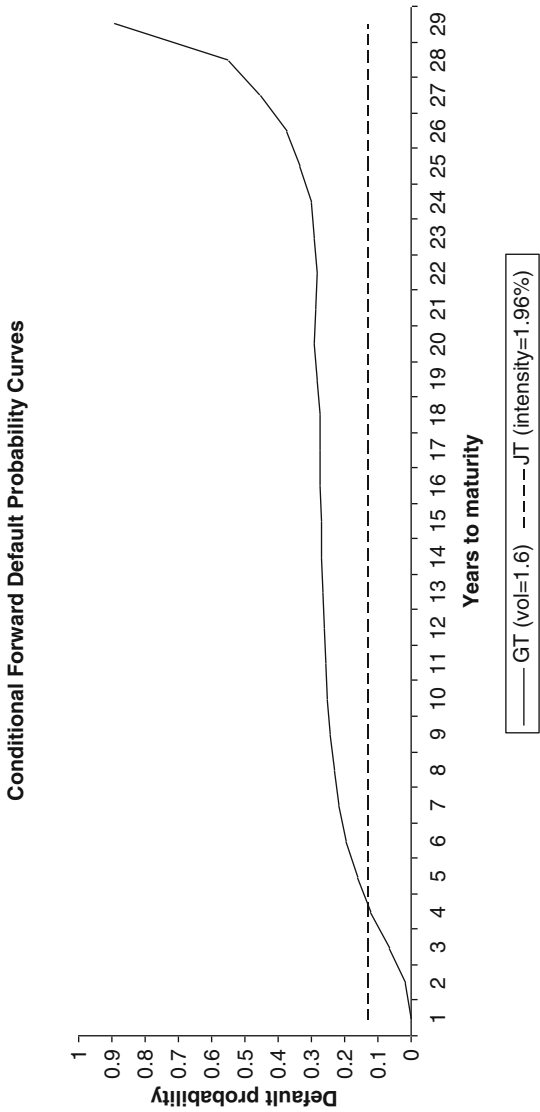
Note: Figure 1 illustrates the survival probability curves under various asset volatility scenarios of the 30-year zero coupon bond under the Geske-Johnson model. The Asset value is set to be \$184. The bond has no coupon and a face value of \$110. The yield curve is flat at 5%

Fig. 1 The GJ model: zero coupon bond (Note: Fig. 1 illustrates the survival probability curves under various asset volatility scenarios of the 30-year zero coupon bond under the Geske-Johnson model. The asset value is set to be \$184. The bond has no coupon and a face value of \$110. The yield curve is flat at 5%)



Note: Figure 1a illustrates the default probability curve. All parameters are identical to those in Figure 1.

Fig. 2 The GJ model: zero coupon bond (Note: Fig. 2 illustrates the default probability curve. All parameters are identical to those in Fig. 1)



Note: Figure 1b illustrates the default probability curves under GJ and JT models for the volatility level of 1.6. All parameters are identical to those in Figure 1.

Fig. 3 Comparison of GJ and JT models: zero coupon bond (Note: Fig. 3 illustrates the default probability curves under GJ and JT models for the volatility level of 1.6. All parameters are identical to those in Fig. 1)

Our next analysis is to keep bond value fixed, so that we can examine the default probability curve with the risk of the bond controlled. We assume that the company issues only one coupon bond at 10%. At the volatility level of 0.4, the asset value is \$123, at 0.6, it is \$290, at 1.0, it is \$15,000. Figures 4 and 5 demonstrate the Geske-Johnson model for various volatility levels but keep the bond at par. We can see that for the same par bond, the default and survival probability curves are drastically different as the asset/volatility combination changes. This is a feature not captured by either the Duffie-Singleton or the Jarrow-Turnbull model.

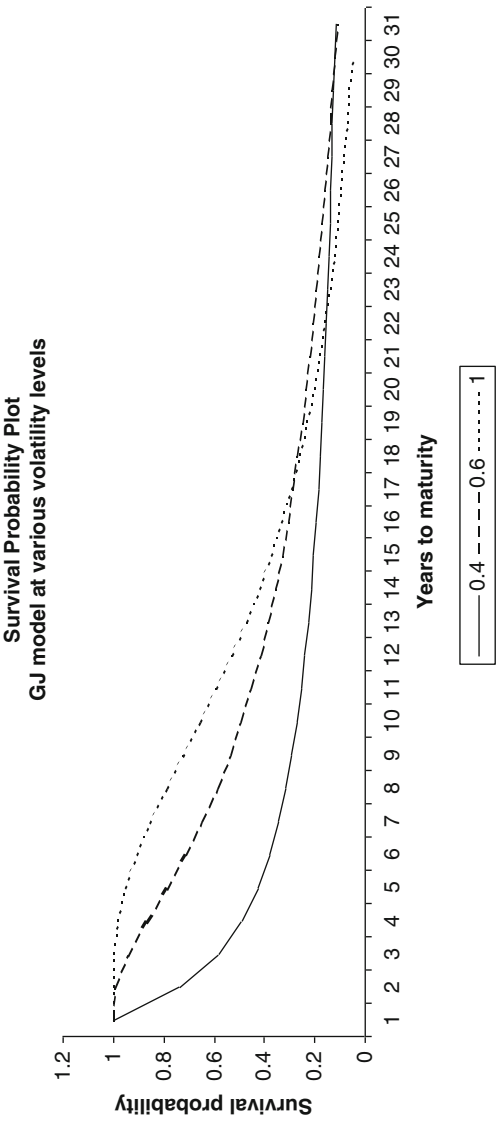
To compare to the Jarrow-Turnbull and Duffie-Turnbull models, we keep the case where the volatility level is 0.6 and asset value is 290. The recover rates of both Jarrow-Turnbull and Duffie-Singleton models are assumed to be 0.4. Figures 6 and 7 show the survival and default probability curves of the three models. The flat conditional forward default probability for the Jarrow-Turnbull model is solved to be 7.60% (or equivalently the intensity rate is 7.74%.) The conditional forward default probabilities for the Duffie-Singleton model are certainly non-constant. The “recovery-adjusted” continuously compounded discount rate is 9.52%. From the survival probability curves (Fig. 6), it is seen that the Geske-Johnson and Jarrow-Turnbull models can be close. But the default probability curves (Fig. 7) demonstrate that the default pattern can be quite different.

75.5 Conclusion

In this chapter, we provide a general framework that brings consistency between the reduced form and structural models. The structural models we consider in this chapter is not those of the “barrier-type” that assumes exogenous barriers but the Geske-Johnson model that allows the default points (strike price) to be endogenously computed. We show that the “true structural” model of Geske-Johnson can be simplified to barrier-type (extended Merton) if the endogenous default points are not internally solved for but exogenously given.

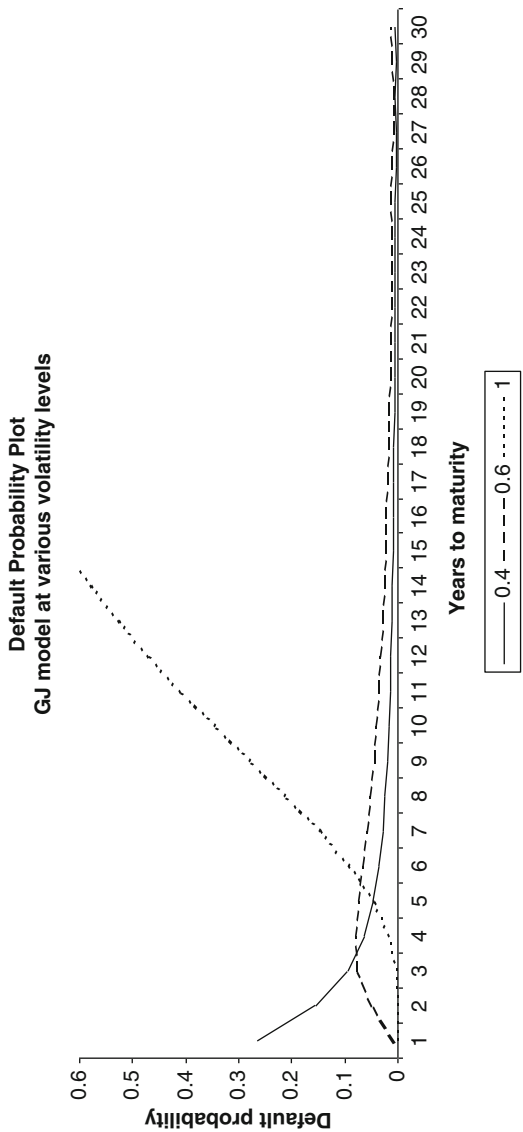
In a discrete time binomial framework, we show that any set of risky cash flows and recovery can be priced by a simple formula. This formula is same for both structural and reduced form models. This formula allows us to compare various models because they only differ in the recovery computation. Different recovery assumptions result in different survival and default probabilities. In calibration, the differences in recovery amounts and in probabilities balance each other out, as we demonstrate, because each model price is made to match the market price. However, these different recovery values and probabilities should result in large differences in derivatives prices. We demonstrate that large differences exist even for the simplest credit derivative contract – credit default swaps.

Finally, In order to compare the “true structural” Geske-Johnson model with the reduced form models represented by Jarrow and Turnbull (1995) and Duffie and Singleton (1997, 1999), we extend the Geske-Johnson model to incorporate random interest rates and a non-flat volatility curve. In a series of appendices, we demonstrate how to implement the Geske-Johnson model with random interest rates and a volatility curve.



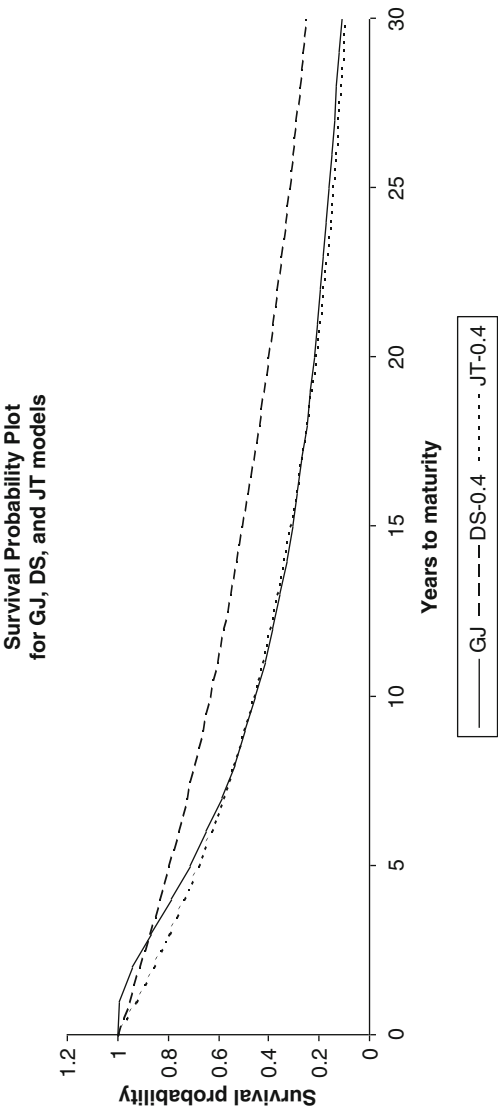
Note: Figure 2 illustrates the survival probability curves under various asset volatility scenarios of the 30-year Geske-Johnson model. The Asset values are 123, 290, 15,000 for volatility levels of 0.4, 0.6, and 1.0 respectively. The bond has a coupon of \$10 and price of par.

Fig. 4 The GJ model: 10% coupon bond (Note: Fig. 4 illustrates the survival probability curves under various asset volatility scenarios of the 30-year Geske-Johnson model. The asset values are 123, 290, 15,000 for volatility levels of 0.4, 0.6, and 1.0 respectively. The bond has a coupon of \$10 and price of par)



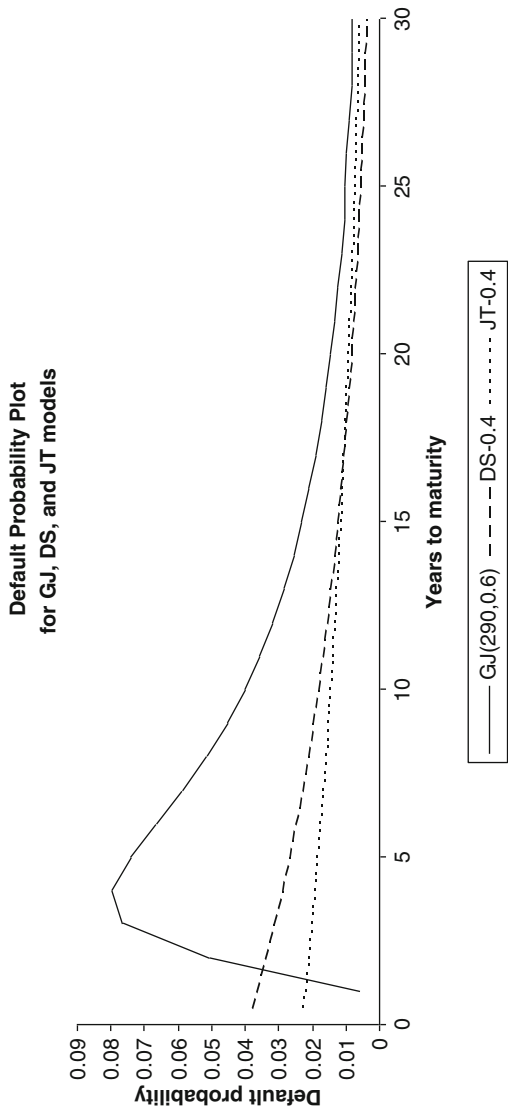
Note: Figure 2a illustrates the default probability curves under various asset volatility scenarios of the 30-year Geske-Johnson model. The Asset values are 123, 290, 15,000 for volatility levels of 0.4, 0.6, and 1.0 respectively. The bond has a coupon of \$10 and price of par. All parameters are identical to those in Figure 2.

Fig. 5 The GJ model: 10% coupon bond (Note: Fig. 5 illustrates the default probability curves under various asset volatility scenarios of the 30-year Geske-Johnson model. The asset values are 123, 290, 15,000 for volatility levels of 0.4, 0.6, and 1.0 respectively. The bond has a coupon of \$10 and price of par. All parameters are identical to those in Fig. 4)



Note: Figure 3 illustrates the survival probability curves under the 30-year Geske-Johnson, Duffie-Singleton with 0.4 recovery ratio, and Jarrow-Turnbull with 0.4 recovery ratio models. The Asset value is set to be \$290 and volatility 0.6 so that the bond is priced at par. The bond has \$10 coupon and a face value of \$110.

Fig. 6 Comparison of the GJ, DS and JT models: par coupon bond (Note: Fig. 6 illustrates the survival probability curves under the 30-year Geske-Johnson, Duffie-Singleton with 0.4 recovery ratio, and Jarrow-Turnbull with 0.4 recovery ratio models. The asset value is set to be \$290 and volatility 0.6 so that the bond is priced at par. The bond has \$10 coupon and a face value of \$110)



Note: Figure 3a illustrates that under a 10% coupon bond, the Geske-Johnson model can generate desired default probability curve. All parameters are identical to those in Figure 3.

Fig. 7 Comparison of the GJ, DS and JT models: par coupon bond (Note: Fig. 7 illustrates that under a 10% coupon bond, the Geske-Johnson model can generate desired default probability curve. All parameters are identical to those in Fig. 6)

Appendix: Derivation of (18)

We shall derive (18) by induction. A T_1 -maturity zero coupon bond with a barrier X_1 is:

$$\begin{aligned} D(0, T_1) &= E_0^Q \left[e^{-\int_0^{T_1} r(u) du} \min \{A(T_1), X_1\} \right] \\ &= P(0, T_1) E_0^{F_1} [\min \{A(T_1), X_1\}] \\ &= A(0) [1 - N(h_1^+(X_1))] + P(0, T_1) X_1 N(h_1^-(X_1)) \end{aligned} \quad (33)$$

where $h_1^\pm(X_1)$ is defined in the text. A T_2 -maturity zero coupon bond with a barrier X_2 has the following value at T_1 :

$$\begin{cases} E_{T_1}^Q \left[e^{-\int_{T_1}^{T_2} r(u) du} \min \{A(T_2), X_2\} \right] & A(T_1) > X_1 \\ 0 & A(T_1) \leq X_1 \end{cases} \quad (34)$$

Hence, it has the following value today:

$$\begin{aligned} D(0, T_2) &= E_0^Q \left[e^{-\int_0^{T_1} r(u) du} E_{T_1}^Q \left[e^{-\int_{T_1}^{T_2} r(u) du} \min \{A(T_2), X_2\} \right] I_{\{A(T_1) > X_1\}} \right] \\ &= E_0^Q \left[E_{T_1}^Q \left[e^{-\int_0^{T_2} r(u) du} (A(T_2) I_{\{A(T_2) < X_2\}} + X_2 I_{\{A(T_2) > X_2\}}) \right] I_{\{A(T_1) > X_1\}} \right] \\ &= P(0, T_2) E_0^{F_n} [A(T_2) I_{\{A(T_2) < X_2 \wedge A(T_1) > X_1\}} + X_2 I_{\{A(T_2) > X_2 \wedge A(T_1) > X_1\}}] \\ &= P(0, T_2) E_0^{F_n} [A(T_2) (I_{\{A(T_1) > X_1\}} - I_{\{A(T_2) > X_2 \wedge A(T_1) > X_1\}}) + X_2 I_{\{A(T_2) > X_2 \wedge A(T_1) > X_1\}}] \\ &= A(0) [N_1^+ - N_2^+] + P(0, 2) X_2 N_2^- \end{aligned} \quad (35)$$

where the second to last line is obtained by appropriately dividing the integration region. Carrying out the expectation using the standard log normal results should yield the result desired. Similar procedure is applied to any arbitrary T_n .

N_n^+ and N_n^-

In this appendix, we show that N_n^+ and N_n^- are both survival probabilities, but under different probability measures. For the ease of composition, we use $N_1^\pm = N(h_1^\pm(X_1))$ where $N(\cdot)$ is a standard univariate normal probability

function, as an example. The general case, while tedious algebraically, is straightforward.

To simplify notation, we drop “1” from $h_1^\pm$, X_1 , T_1 , and measure F_1 . Note that:

$$\begin{aligned} N(h^+) &= \frac{P(0, T)}{A(0)} E_0^F [A(T) I_{\{A(T) > X\}}] \\ N(h^-) &= E_0^F [I_{\{A(T) > X\}}] \end{aligned} \quad (36)$$

Equation 36 can be written as:

$$\begin{aligned} E_0^F [A(T) I_{\{A(T) > X\}}] &= E_0^F [A(T)] E_0^{F*} [I_{\{A(T) > X\}}] \\ &= \frac{A(0)}{P(0, T)} E_0^{F*} [I_{\{A(T) > X\}}] \end{aligned} \quad (37)$$

The change of measure from F to F* is done as follows:

$$\frac{dF^*}{dF} = \frac{A(T)}{E_0^F [A(T)]} \quad (38)$$

Given that A is a log normal process, we immediately obtain that the above derivative as a change of measure of $\sqrt{\text{var}[\ln(A(T))]}$ under the F measure. In the following, we sketch the basic math of forward measure.

Assume an interest rate process under the Q measure generally as:

$$dr(t) = \alpha(r, t)dt + \sigma(r, t)dW^Q(t) \quad (39)$$

Then, by (3), we have the Radon-Nykodym derivative defined as follows:

$$\eta(t, T) = \frac{\Lambda(t, T)}{P(t, T)} \quad (40)$$

Applying Ito's lemma on the bond price:

$$\begin{aligned} 0 &= \ln P(T, T) \\ &= \ln P(t, T) + \int_t^T \frac{1}{P(u, T)} \left[P_u(u, T)du + P_r(u, T)dr + \frac{1}{2} P_{rr}(u, T)(dr)^2 \right] dW^Q(u) \\ &\quad - \int_t^T \frac{1}{2} \left[\frac{\sigma(r, u)P_r(u, T)}{P(u, T)} \right]^2 du \end{aligned}$$

$$\begin{aligned}
&= \ln P(t, T) + \int_t^T \frac{1}{P(u, T)} [P_u(u, T) du + P_r(u, T) \widehat{\mu}(r, u) \\
&\quad + \frac{1}{2} P_{rr}(u, T) \sigma(r, u)^2] du + \int_t^T \frac{1}{P(u, T)} P_r(u, T) \sigma(r, u) dW^Q(u) \\
&\quad - \int_t^T \frac{1}{2} \left[\frac{\sigma(r, u) P_r(u, T)}{P(u, T)} \right]^2 du \quad (41) \\
&= \ln P(t, T) + \int_t^T r(u) du + \int_t^T \frac{1}{P(u, T)} P_r(u, T) \sigma(r, u) dW^Q(u) \\
&\quad - \int_t^T \frac{1}{2} \left[\frac{\sigma(r, u) P_r(u, T)}{P(u, T)} \right]^2 du
\end{aligned}$$

Letting:

$$\theta(t, T) = \frac{\sigma(r, t) P_r(t, T)}{P(t, T)} \quad (42)$$

and moving the first two terms to the left:

$$\begin{aligned}
&-\int_t^T r(u) du - \ln P(t, T) = \int_t^T \theta(u, T) dW^Q(u) - \int_t^T \frac{1}{2} \theta(u, T)^2 du \\
&\frac{\Lambda(t, T)}{P(t, T)} = \eta(t, T) = \exp \left(\int_t^T \theta(u, T) dW^Q(u) - \int_t^T \frac{1}{2} \theta(u, T)^2 du \right) \quad (43)
\end{aligned}$$

This implies the Girsanov transformation of the following:

$$W^Q(t) = W^{F_T}(t) + \int_t^T \sigma(r, u) \frac{P_r(u, T)}{P(u, T)} du \quad (44)$$

The interest rate process under the forward measure henceforth becomes:

$$dr(t) = \left[\alpha(r, t) + \sigma(r, t)^2 \frac{P_r(t, T)}{P(t, T)} \right] dt + \sigma(r, t) dW^{F_T}(t) \quad (45)$$

Note that the forward measure derived above is quite general. It does not depend on any specific assumption on the interest rate process. In the following, “random

equity and random interest rate,” we do need normally distributed interest rates, or there is no solution to option pricing.

$$\begin{aligned} \begin{bmatrix} \frac{dA(t)}{A(t)} \\ dr(t) \end{bmatrix} &= \begin{bmatrix} r(t) \\ \alpha(r, t) \end{bmatrix} dt + \begin{bmatrix} \sigma_A & 0 \\ 0 & \sigma_r(r, t) \end{bmatrix} \\ &\quad \times \begin{bmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} dW_S^Q(t) \\ dW_r^Q(t) \end{bmatrix} \end{aligned} \quad (46)$$

Using the forward measure, we get:

$$\begin{aligned} \begin{bmatrix} \frac{dA(t)}{A(t)} \\ dr(t) \end{bmatrix} &= \begin{bmatrix} r(t) + \rho\sigma_A\sigma_r(r, t) \frac{P_r(t, T)}{P(t, T)} \\ \alpha(r, t) + \sigma_r(r, t)^2 \frac{P_r(t, T)}{P(t, T)} \end{bmatrix} dt \\ &\quad + \begin{bmatrix} \sigma_A & 0 \\ 0 & \sigma_r(r, t) \end{bmatrix} \times \begin{bmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} dW_S^Q(t) \\ dW_r^{F_T}(t) \end{bmatrix} \end{aligned} \quad (47)$$

$$\begin{aligned} A(T) &= A(t) \exp \left(\int_t^T \left[r(u) + \rho\sigma_A\sigma_r(r, t) \frac{P_r(u, T)}{P(u, T)} \right] du \right. \\ &\quad \left. + \int_t^T \sigma dW^{F_T}(u) - \frac{1}{2} \int_t^T \sigma^2 du \right) \end{aligned} \quad (48)$$

where

$$dW^{F_T}(t) = \sqrt{1-\rho^2} dW_S^Q(t) + \rho dW_r^{F_T}(t)$$

From the result of Ito's lemma, we can then arrive at the following result:

$$\begin{aligned} A(T) &= \frac{A(t)}{P(t, T)} \exp \left(\int_t^T \left[-\frac{1}{2} \left[\sigma_r(r, t) \frac{P_r(u, T)}{P(u, T)} \right]^2 \right. \right. \\ &\quad \left. \left. + \rho\sigma_A\sigma_r(r, t) \frac{P_r(u, T)}{P(u, T)} - \frac{1}{2} \sigma^2 \right] du \right. \\ &\quad \left. - \int_t^T \sigma_r(r, t) \frac{P_r(u, T)}{P(u, T)} dW_r^{F_T}(u) + \int_t^T \sigma_A dW^{F_T}(u) \right) \end{aligned} \quad (49)$$

Hence,

$$\begin{aligned} E_t^{F_T}[\ln S(T)] &= \ln S(t) - \ln P(t, T) + \frac{1}{2}v(t, T)^2 \\ \text{var}_t^{F_T}[\ln S(T)] &= v(t, T)^2 \end{aligned} \quad (50)$$

where

$$v(t, T)^2 = \int_t^T \left[\left[\sigma_r(r, t) \frac{P_r(u, T)}{P(u, T)} \right]^2 - 2\rho\sigma_A\sigma_r(r, t) \frac{P_r(u, T)}{P(u, T)} + \sigma_A^2 \right] du$$

Derivation of (24)

The Geske-Johnson model is usually written as the multi-variate normal probability format. As a result, the final solution looks more complex than it really is. Once, we understand the recursive relationship in zero coupon bond formulas, the final result is very straightforward to recognize. Hence, in this appendix, we provide a simple three period example to demonstrate how we can easily derive and streamline the Geske-Johnson model. The rest can be obtained by induction.

For the 1-year zero, it is identical to the Merton-Rabinovitch model:

$$\begin{aligned} D(0, T_1) &= V_{\text{GJ}}(0, T_1) = P(0, T_1)K_{11} [N_1(h_1^-(K_{11})) - 0] \\ &\quad + A(0)[1 - N_1(h_1^+(K_{11}))] \\ &= P(0, T_1)K_{11}\Pi_1^-(K_{11}) + A(0)[1 - \Pi_1^+(K_{11})] \end{aligned} \quad (51)$$

Note that $K_{11} = X_1$, the first coupon. The second line of the above equation is merely a notation change. The simpler notation allows to more easily label high dimension normal probabilities with different strikes.

For the 2-year zero, we need to solve for an internal solution K_{12} that equates

$$A(T_1) = D(T_1, T_2) + X_1 \quad (52)$$

where

$$\begin{aligned} D(T_1, T_2) &= E_{T_1}^Q[\Lambda(T_1, T_2) \min\{A(T_2), X_2\}] \\ &= P(T_1, T_2)N(d_-) + A(T_1)[1 - N(d_+)] \end{aligned} \quad (53)$$

and

$$d_{\pm} = \frac{\ln A(T_1) - \ln P(T_1, T_2) - \ln X_2 \pm \frac{1}{2}v^2(T_1, T_2)}{v(T_1, T_2)} \quad (54)$$

is a Merton-Rabinovitch result again. Note that $v^2(\cdot, \cdot)$ is defined in (18). We should note that since $P(T_1, T_2)$ is a function of the interest rate at time T_1 , i.e. $r(T_1)$. As a result, K_{12} is also a function of $r(T_1)$. For the convenience of later derivation, we rewrite (66) in its original integral form as follows. Also for notational convenience, we shorten the following notation: $D(T_i, T_j) = D_{ij}$, $P(T_i, T_j) = P_{ij}$, $A(T_i) = A_i$ and $r(T_i) = r_i$.

$$D_{12} = \int_{-\infty}^{\infty} \int_{X_2}^{\infty} \Lambda_{12} X_2 \varphi(A_2, r_2) dA_2 dr_2 + \int_{-\infty}^{\infty} \int_0^{X_2} A_{12} A_2 \varphi(A_2, r_2) dA_2 dr_2 \quad (53a)$$

This expression allows us to easily integrate with other integrals. The value of the 2-year zero price at T_1 is:

$$\begin{cases} D_{12} & A_1 > K_{12} \\ A_1 - X_1 & X_1 < A_1 \leq K_{12} \\ 0 & A_1 \leq X_1 \end{cases} \quad (55)$$

To obtain the current value of the 2-year zero, we simply integrate these payoffs at its corresponding region. Note that K_{12} is a function of the interest rate.

$$\begin{aligned} D(0, 2) = & \underbrace{\int_{-\infty}^{\infty} \int_{K_{12}}^{\infty} \Lambda_{0,1} E[\Lambda_{1,2} \min \{A_2, F_2\}] \varphi(A_1, r_1) dA_1 dr_1}_{\Delta_1} \\ & + \underbrace{\int_{-\infty}^{\infty} \int_{X_1}^{K_{12}} \Lambda_{0,1} (A_1 - F_1) \varphi(A_1, r_1) dA_1 dr_1}_{\Delta_2} \end{aligned} \quad (56)$$

The second term can be shown as:

$$\begin{aligned} \Delta_2 = & \int_{-\infty}^{\infty} \int_{K_{11}}^{\infty} \Lambda_{01} (V_1 - F_1) \varphi(V_1, r_1) dV_1 dr_1 \\ & - \int_{-\infty}^{\infty} \int_{K_{12}}^{\infty} \Lambda_{01} (V_1 - F_1) \varphi(V_1, r_1) dV_1 dr_1 \\ = & V_0 [\Pi_1^+(K_{11}) - \Pi_1^+(K_{12})] - P_{01} X_1 [\Pi_1^-(K_{11}) - \Pi_1^-(K_{12})] \end{aligned} \quad (57)$$

The first term can be valued as:

$$\begin{aligned}
\Delta_1 &= \int_{-\infty}^{\infty} \int_{K_{12}}^{\infty} \Lambda_{01} \left[\int_{-\infty}^{\infty} \int_0^{K_{22}} \Lambda_{12} A_2 \varphi(A_2, r_2 | A_1, r_1) dA_2 dr_2 \right] \\
&\quad \varphi(A_1, r_1) dA_1 dr_1 \\
&\quad + \int_{-\infty}^{\infty} \int_{K_{12}}^{\infty} \Lambda_{01} \left[\int_{-\infty}^{\infty} \int_{K_{22}}^{\infty} \Lambda_{12} X_2 \varphi(A_2, r_2 | A_1, r_1) dA_2 dr_2 \right] \\
&\quad \varphi(A_1, r_1) dA_1 dr_1 \\
&= \int_{-\infty}^{\infty} \int_{K_{12}}^{\infty} \Lambda_{0,1} \left[\int_{-\infty}^{\infty} \int_0^{\infty} \Lambda_{12} A_2 \varphi(A_2, r_2 | A_1, r_1) dA_2 dr_2 \right] \\
&\quad \varphi(A_1, r_1) dA_1 dr_1 \\
&\quad - \int_{-\infty}^{\infty} \int_{K_{12}}^{\infty} \Lambda_{01} \left[\int_{-\infty}^{\infty} \int_{K_{22}}^{\infty} \Lambda_{12} A_2 \varphi(A_2, r_2 | A_1, r_1) dA_2 dr_2 \right] \\
&\quad \varphi(A_1, r_1) dA_1 dr_1 \\
&\quad + \int_{-\infty}^{\infty} \int_{K_{12}}^{\infty} \Lambda_{01} \left[\int_{-\infty}^{\infty} \int_{K_{22}}^{\infty} \Lambda_{12} X_2 \varphi(A_2, r_2 | A_1, r_1) dA_2 dr_2 \right] \\
&\quad \varphi(A_1, r_1) dA_1 dr_1 \\
&= A_0 \Pi_1^+(K_{12}) - A_0 \Pi_2^+(K_{12}, K_{22}) + P_{02} X_2 \Pi_2^-(K_{12}, K_{22})
\end{aligned} \tag{58}$$

Hence, the 2-year coupon bond, returning to the original notation, is:

$$\begin{aligned}
V_{\text{GJ}}(0, T_2) &= D(0, T_1) + D(0, T_2) \\
&= D(0, T_1) + \Delta_1 + \Delta_2 \\
&= P(0, T_1) X_1 \Pi_1^-(X_1) + P(0, T_2) X_2 \Pi_2^-(K_{12}, X_2) \\
&\quad + A_0 [1 - \Pi_2^+(K_{12}, X_2)]
\end{aligned} \tag{59}$$

where

$$\begin{aligned}
\Pi_1^{\pm}(K) &= \int_{-\infty}^{\infty} \int_K^{\infty} \varphi^{\pm}(A_1, r_1) dA_1 dr_1 \\
\Pi_2^{\pm}(K_1, K_2) &= \int_{-\infty}^{\infty} \int_{K_1}^{\infty} \int_{K_2}^{\infty} \varphi^{\pm}(A_1, A_2, r_1) dA_2 dA_1 dr_1
\end{aligned} \tag{60}$$

It is shown in Appendix A that the last two integrals in the above equation can be written as:

$$\int_K^\infty A_1 \varphi^-(A_1|r_1) dA_1 = A_0 \int_K^\infty \varphi^+(A_1|r_1) dA_1 \quad (61)$$

and

$$\int_{K_1}^\infty \int_{K_2}^\infty A_2 \varphi^-(A_1, A_2|r_1) dA_2 dA_1 = A_0 \int_{K_1}^\infty \int_{K_2}^\infty \varphi^+(A_1, A_2|r_1) dA_2 dA_1 \quad (62)$$

where the density is adjusted by the volatility.

Following the same procedure, although tedious, we can derive the 3-year zero as:

$$\begin{aligned} D(0, T_3) = & P(0, T_1) X_1 [\Pi_1^-(K_{13}) - \Pi_1^-(K_{12})] \\ & + P(0, T_2) X_2 [\Pi_2^-(K_{13}, K_{23}) - \Pi_2^-(K_{12}, K_{22})] \\ & + P(0, T_3) X_3 \Pi_3^-(K_{13}, K_{23}, K_{33}) \\ & + A(0) [\Pi_2^+(K_{12}, K_{22}) - \Pi_3^+(K_{13}, K_{23}, K_{33})] \end{aligned} \quad (63)$$

Now, we should observe a pattern for the zero coupon bond prices, which gives (24).

Implementation of (24)

The closed form Geske-Johnson model (with constant volatility and constant interest rates) can be computed efficiently only when $n \leq 2$. When $n > 2$, then the multivariate normal probability functions can not be implemented efficiently, particularly in high dimensions. We use the standard equity binomial model with various payoffs to pick up the survival probabilities, zero bond values, and the equity (compound option) value.¹⁶ Note that in the binomial model, there is no need to solve for the implied strike price, K_{ij} . The (compound) option value is directly computed off the actual strikes, X_i .

In (24), we have both volatility and interest rate to be non-constant. We allow volatility to be a deterministic function of time (for calibration) and interest rates to be random (to capture interest rate risk). We shall sketch briefly in this appendix how (24) is implemented.

¹⁶We learned via private conversation that the implementation of the Geske-Johnson model in Eom et al. (2002) is computed using the binomial model.

First, the deterministic volatility function is handled by changing time interval, suggested by Amin (1991). In the binomial model, the up and down sizes are determined by:

$$\begin{aligned} u &= e^{\sigma\sqrt{\Delta t}} \\ d &= e^{-\sigma\sqrt{\Delta t}} \end{aligned} \quad (64)$$

If the volatility is changing over time, we simply adjust Δt so that $\sigma\sqrt{\Delta t}$ is constant. In this case, both u and d are constants over time and the tree recombines.

To incorporate stochastic interest rates, we first build a lattice for a “risk-neutral” bivariate Brownian motion process:

$$\begin{aligned} \begin{bmatrix} d \ln A \\ dr \end{bmatrix} &= \begin{bmatrix} r - \sigma^2/2 \\ \alpha(\mu - r) \end{bmatrix} dt + \begin{bmatrix} \sigma & 0 \\ 0 & \delta \end{bmatrix} \\ &\quad \times \begin{bmatrix} \sqrt{1 - \rho^2} & \rho \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} dW_1 \\ dW_2 \end{bmatrix} \end{aligned} \quad (65)$$

where $E[dW_1 dW_2] = 0$. We then set up a binomial lattice as follows:

$$\begin{array}{c} \sqrt{\Delta t} \\ 0 \\ -\sqrt{\Delta t} \end{array}$$

We label asset value and interest rate at various nodes as follows:

$$\begin{array}{ccc} & & < \ln A_{22}, r_{22} > \\ & < \ln A_{11}, r_{11} > & \\ < \ln A_{00}, r_{00} > & < \ln A_{21}, r_{21} > \\ & < \ln A_{10}, r_{10} > & \\ & & < \ln A_{20}, r_{20} > \end{array}$$

Then we first show that the asset values recombine. Given that:

$$\begin{aligned} \ln A_{11} &= \ln A_0 + (r_0 - \sigma^2/2)\Delta t + \sigma\sqrt{\Delta t} \\ \ln A_{10} &= \ln A_0 + (r_0 - \sigma^2/2)\Delta t - \sigma\sqrt{\Delta t} \end{aligned} \quad (66)$$

It is straightforward that:

$$\begin{aligned} \ln A_{21} &= \ln A_{10} + (r_0 - \sigma^2/2)\Delta t + \sigma\sqrt{\Delta t} \\ &= \ln A_{11} + (r_0 - \sigma^2/2)\Delta t - \sigma\sqrt{\Delta t} \end{aligned} \quad (67)$$

We now show that the interest rate lattice recombines approximately. Note that:

$$\begin{aligned} r_{11} &= r_0(1 - \alpha\Delta t) + \alpha\mu\Delta t + \delta\sqrt{\Delta t} \\ r_{10} &= r_0(1 - \alpha\Delta t) + \alpha\mu\Delta t - \delta\sqrt{\Delta t} \end{aligned} \quad (68)$$

Hence,

$$\begin{aligned} r_{21} &= r_{10}(1 - \alpha\Delta t) + \alpha\mu\Delta t + \delta\sqrt{\Delta t} \\ &= \left[r_0(1 - \alpha\Delta t) + \alpha\mu\Delta t - \delta\sqrt{\Delta t} \right] (1 - \alpha\Delta t) \\ &\quad + \alpha\mu\Delta t + \delta\sqrt{\Delta t} \\ &= r_0(1 - 2\alpha\Delta t) + 2\alpha\mu\Delta t \\ &= r_{11}(1 - \alpha\Delta t) + \alpha\mu\Delta t - \delta\sqrt{\Delta t} \end{aligned} \quad (69)$$

This result is an approximated one because higher order terms are ignored.¹⁷

Closed Form Solution for $v(0, T_i)$: Rabinovitch

Following the interest rate model defined in Appendix D (i.e. the Vasicek model), the zero coupon bond price satisfies the following closed form equation (Rabinovitch 1989):

$$P(0, t) = e^{-rF(0, t) - G(0, t)} \quad (70)$$

where

$$\begin{aligned} F(0, t) &= \frac{1 - e^{-\alpha t}}{\alpha} \\ G(0, t) &= (t - F(0, t)) \left(\mu - \frac{\delta^2}{2\alpha} \right) + \frac{\delta^2 F(0, t)^2}{4\alpha} \end{aligned}$$

Hence, by Ito's lemma we have,

$$\sigma_P^2(t, T_i) = \left[\frac{\partial P(t, T_i)}{\partial r} \delta \right]^2 = F(t, T_i)^2 \delta^2 \quad (71)$$

and

¹⁷ If higher order terms are not ignored, then the lattice does not combine. In the case where the lattice does not recombine, we take the average of two non-recombining values at each node and proceed. We discover that this method produces extremely close results when we compare the zero coupon bond price against the Vasicek closed form model. We use the same technique with both deterministic volatility function and random interest rate are used. Note that in here, since Δt is different period by period, the interest rate dimension of the lattice does not recombine.

$$\int_0^t \sigma_p^2(u, T_i) du = \int_0^t F(u, T_i)^2 \delta^2 du = \delta^2 t + \left((t - 2F(0, t)) + \frac{1 - e^{-2\alpha t}}{2\alpha} \right) \left(\frac{\delta}{\alpha} \right)^2 \quad (72)$$

References

- Amin, Kausik. 1991. On the computation of continuous time option prices using discrete approximations. *Journal of Financial and Quantitative Analysis* 26 (4): 477–496. December.
- Anderson, Ronald, and Suresh Sundaresan. 1996. Design and valuation of debt contracts. *Review of Financial Studies* 9: 37–68.
- Bélanger, Alain, Steven Shreve, and Dennis Wong. 2002. *A unified model for credit derivatives, Working paper*. Carnegie Mellon University.
- Black, Fischer, and John Cox. 1976. Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance* 31: 351–367.
- Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–659.
- Cox, John, Jonathan Ingersoll, and Stephen Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53: 385–408.
- Duffie, Darrell and David Lando. 1997. Term structure of credit spreads with incomplete accounting information. Forthcoming in *Econometrica*.
- Duffie, Darrell, and Kenneth J. Singleton. 1997. An econometric model of the term structure of interest-rate swap yields. *Journal of Finance* 52 (4): 1287–1321. September.
- . 1999. Modeling the term structure of defaultable bonds. *Review of Financial Studies* 12: 687–720.
- Eom, Young Ho, Jean Helwedge, and Jing-zhi Huang. 2002. Structural models of corporate bond pricing: An empirical analysis. *Review of Financial Studies*, forthcoming.
- Geske, Robert. 1977. The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis*: 541–552.
- Geske, Robert, and Herb Johnson. 1984. The valuation of corporate liabilities as compound options: A correction. *Journal of Financial and Quantitative Analysis* 19 (2): 231–232.
- Ho, Thomas S.Y., and Sang-bin Lee. 1986. Term structure movements and pricing interest rate contingent claims. *Journal of Finance* 41 (5): 1011–1029. December.
- Hull, John. 2000. *Options, futures, and other derivative contracts*. Prentice Hall.
- Hull, John, and Alan White. 1990. Pricing interest-rate-derivative securities. *Review of Financial Studies* 3 (4): 573–592.
- Jamshidian, Farshid. 1987. *Pricing of contingent claims in the one factor term structure model*, Working paper, Merrill Lynch Capital Markets.
- Jarrow, Robert, and Stuart Turnbull. 1995. Pricing derivatives on financial securities subject to default risk. *Journal of Finance* 50: 53–86.
- Jarrow, Robert A., and Fan Yu. 2001. Counterparty risk and the pricing of defaultable securities. *Journal of Finance* 56 (5): 1765–1799. October.
- Jarrow, Robert, David Lando, and Stuart Turnbull. 1997. A Markov model for the term structure of credit spreads. *Review of Financial Studies* 10: 481–523.
- Lando, David. 1998. On Cox processes and credit risky securities. *Review of Derivatives Research* 2 (2–3): 99–120. December.
- Leland, Hayne. 1994. Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance* 49 (4): 1213–1252.

- Leland, Hayne, and Klaus Toft. 1996. Optimal capital structure, endogenous Bankruptcy, and the term structure of credit spreads. *Journal of Finance* 51 (3): 987–1019.
- Longstaff, Francis, and Eduardo Schwartz. 1995. Valuing risky debt: A new approach. *Journal of Finance*: 789–820.
- Madan, Dilip B., and Haluk Unal. 2000. A two-factor hazard rate model for pricing risky debt and the term structure of credit spreads. *Journal of Financial & Quantitative Analysis* 35 (1): 43–65. March.
- Merton, Robert C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29: 449–470.
- Rabinovitch, R..1989. Pricing stock and bond options when the default free rate is stochastic. *Journal of Financial and Quantitative Analysis*.
- Schönbucher, Philipp J. 1998. Term structure modelling of defaultable bonds. *Review of Derivatives Research* 2 (2–3): 161–192. December.
- Uhrig-Homburg, Marliese. 2002. Valuation of defaultable claims: A survey. *Schmalenbach Business Review* 54: 24–57. January.
- Vasicek, O. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5: 177–188.
- Zhou, Chunsheng. 2001. An analysis of default correlations and multiple defaults. *Review of Financial Studies* 14 (2): 555–576.

Bankruptcy Prediction Studies Across Countries Using Multiple Criteria Linear Programming (MCLP) and Other Data Mining Approaches

76

Wikil Kwak and Yong Shi

Abstract

In this study, we want to show the usefulness of Multiple Criteria Linear Programming (MCLP) data mining approach to predict bankruptcy across countries such as Japan, Korea, and China compared with those of US studies. Bankruptcy studies are interesting to investors, managers, regulators, and other external users, but traditionally we used logit or probit types of statistical methods to predict firm bankruptcy. Recently, data mining has proven to be a useful tool in accounting and finance for various prediction and classification studies. Many applications have been used in accounting and finance areas where accountants deal with large amounts of computerized operational and financial databases. Eighty-two percent of these studies are predictive studies based on recent survey (Amani, F. A. and A. M. Fadlalla, *International Journal of Accounting Information Systems* 24:32–58, 2017). Segal (*Economics and Business Review* 2:45–64, 2016) proposes more applications of data mining in fraud detection and prevention as advanced technologies in the current information environment. There are higher potential payoffs for data mining applications in these areas. The purpose of this study is to propose an MCLP data mining approach as comparable to other data mining methods for bankruptcy prediction. This approach has proven to be robust and powerful even for a large sample size using a huge financial database. The results of the MCLP approach in a bankruptcy prediction study are promising as this approach performs better than traditional multiple discriminant analysis or logit

W. Kwak (✉)

School of Accounting, University of Nebraska at Omaha, Omaha, NE, USA

e-mail: wkwak@unomaha.edu

Y. Shi

Information Systems and Quantitative Analysis, University of Nebraska at Omaha, Omaha, NE, USA

Chinese Academy of Sciences, Beijing, China

e-mail: yshi@unomaha.edu

analysis using financial data. Similar approaches can be applied to other accounting areas such as auditor changes, internal control weakness prediction, hedging, forecasting stock performance, detection of tax evasion, and an audit-planning tool using financial variables (see Kwak, W., S. Eldridge, S. Yong, and K. Gang, *Journal of Applied Business Research* 25:105–118, 2009, Kwak, W., S. Eldridge, S. Yong, and K. Gang, *Journal of Applied Business Research* 27:73–84, 2011, and Sam, R., J. Osleeb, and K. Si, *International Regional Science Review* 39:77, 2016 for examples).

Keywords

Bankruptcy · Data Mining · Japan · Multiple Criteria Linear Programming

JEL Classification

G33 · M41

76.1 Introduction

Data mining tools are widely used in business applications such as credit card default prediction, a customer's buying behavior analysis, product promotion analysis, and online sales support as more online business are booming because of the pandemic. Recently, more applications in finance and accounting such as bankruptcy prediction, fraud detection, forecasting stock returns, auditor change prediction, and internal control weakness prediction have increased as accounting deals with huge amount of financial data and computational costs are cheaper.

The purpose of this study is to introduce a multiple criteria linear programming (MCLP) approach and other data mining approach for bankruptcy prediction across countries from our previous studies (Kwak et al. 2006a, b, 2012, 2014). Data mining is a tool that can be used to analyze data to generate descriptive or predictive models to solve business problems (Berry and Linoff 1997; Chan and Lewis 2002; Witten and Frank 2000). According to Shi et al. (2002), data mining also can be classified based on its process, methodology, and mathematical tools. From the process view, data mining consists of the following stages: (1) cleaning, (2) integration, (3) selection, (4) transformation (5) data mining, and (6) pattern evaluation or interpreting, see Han and Kamber (2001). Data cleaning is required to remove noise and inconsistent data. Data integration may be needed where multiple data sources are used. Selecting and transforming data for achieving business goals is known as data warehousing. Pattern evaluation or interpretation provides the analysis of mined data for use in business decision making by managers. From the methodological view, data mining can be achieved by association, classification, clustering, and predictions. From the mathematical point of view, the algorithms of data mining can be implemented by binary, induction decision tree, statistics, neural networks, etc. (Shi et al. 2002).

Mousavi and Ouenniche (2018) used statistical models using macroeconomic indicators to predict financial distress using UK data. In contrast, Support Vector Machine methods are found to be highly potential in predicting bankruptcy studies using Korean data (Kim et al. 2018). Klepac and Hampel (2018) also suggested that decision tree approaches are better than support vector machine methods in predicting bankruptcy using EU data. Recently, Batena and Asghari (2020) used logit and genetic algorithm (GA) models to predict bankruptcy and found GA methods to be superior. However, MCLP approach can be comparable to other data mining methods as we have been using these methods in bankruptcy studies across countries.

The challenges may come from creating optimal cost-benefit trade-off mapping between the benefits of quantifying or discrediting financial data and the cost of information loss in the process of granulating financial data for detecting bankruptcy. In computer security detection, the discretization problem may be less severe because of the binary nature of a computer virus or security breach that the detectors deal with, whereas financial data are decimal.

On the other hand, as of now, an MCLP data mining approach has been recently applied to credit card portfolio management. This approach has proven to be robust and powerful for even a large sample size using a huge financial database. The results of the MCLP approach in our bankruptcy prediction study are promising as this approach performs better than traditional multiple discriminant analysis or logistic regression analysis using financial data.

Our study is organized as follows: The next section presents the background of bankruptcy studies. The third section presents an MCLP data mining model. The fourth section describes data collection procedures and reports our empirical results of different countries. The last section concludes with a summary of our findings and future research avenues.

76.2 Background

Altman (1968) proposed a seminal work in the bankruptcy prediction study and used multiple discriminant analysis (MDA) by using five financial ratios to predict a bankruptcy model. Altman et al. (1977) later proposed the ZETA model, but their assumptions of data being normally distributed can be a problem when applying this model. Beaver (1966) used cash flow variables to predict bankruptcy and later studies compared financial ratios with cash flow variables and confirmed that cash flow variables are also useful to predict bankruptcy.

Ohlson (1980) used a logit model, which does not require any assumptions about the prior probability of bankruptcy or the distribution of predictor variables, see Nanda and Pendharkar (2001). The assumptions about the knowledge of prior probabilities and group distributions make both of the above techniques very restrictive in bankruptcy prediction studies. In addition, hold-out sample tests are potentially upwardly biased according to Grice and Dugan (2001) and Platt and Platt

(1990), the differences in the macroeconomic factors are sensitive to specific time periods. Typical finance and accounting bankruptcy studies use financial ratios and other macroeconomic variables to predict bankruptcy and the advantage of these approaches is finding out which variable has significant impact to a firm's bankruptcy. However, if we focus on prediction accuracy rates, data mining methods could be more attractive. Especially the Multiple Criteria Linear Programming data mining method is a promising method compared with other data mining methods.

Korol (2020) proposed long-term trajectories of bankrupt firms using traditional multivariate discriminant analysis and logit models and find that these models are still valid to evaluate long-term growth and early bankruptcy risk beforehand using 17 years of data. Mousavi and Ouenniche (2018) proposed that statistical models using macroeconomic indicators to predict financial distress using 2005–2014 UK data show that their revised statistical methods are still valid. In contrast, Support Vector Machine (SVM) methods are found to be higher potential to predicting bankruptcy studies using 2000–2013 Korean data (Kim et al. 2018). Horak et al. (2020) also proposed that the SVM method show a relatively high accuracy with Czech Republic data. Klepac and Hampel (2018), however, suggested that decision tree approaches are better than support vector machine methods to predict bankruptcy using 2011–2013 EU data. Recently, Batena and Asghari (2020) used logit and genetic algorithm (GA) models to predict bankruptcy and found GA methods to be superior.

The focus of our study is the bankruptcy prediction accuracy rate using past financial data and other variables as other previous studies have used these approaches. Therefore, we propose the MCLP method as comparable to other data mining methods.

76.3 Models of Multiple Criteria Linear Programming Classification

From the aspect of mathematical tools of data mining, the algorithms of data mining can be implemented by many different types of mathematical techniques. For example, Shi et al. (2002) state classification or prediction methods can be constructed by decision trees, statistics, and neural networks as previous bankruptcy studies have used.

Based on Freed and Glover's (1986) linear programming (LP) model, Kou et al. (2003) developed a general problem of data classification by using MCLP. The basic concept of the formulation can be shown as follows:

Given a set of r variables about the bankrupt or non-bankrupt sample firms in the database $\mathbf{a} = (a_1, \dots, a_r)$, let $\mathbf{A}_i = (A_{i1}, \dots, A_{ir}) \in \mathbf{R}^r$ be the sample observations of data for the variables, where $i = 1, \dots, n$ and n is the sample size. We want to determine the coefficients of the variables, denoted by $\mathbf{X} = (x_1, \dots, x_r)^T$, and a boundary value of b to separate two classes: B(Bankrupt) and N (Non-bankrupt); that is,

(a) $A_i X \leq b, A_i \in B$ (Bankrupt) and $A_i X > b, A_i \in N$ (Non-bankrupt).

Now consider two kinds of measurements for better separation of Bankrupt and Non-bankrupt firms. Let α_i be the overlapping degree with respect to A_i , and β_i be the distance from A_i to its adjusted boundary. We also define α to be the maximum overlapping of a two-class boundary for all cases A_i ($\alpha_i < \alpha$) and β to be the minimum distance for all cases A_i from its adjusted boundary ($\beta_i > \beta$).

We use ($\star$) to represent A_i of Non-bankrupt and ($\bullet$) to represent A_i of Bankrupt, then Fig. 1 shows that our goal is to minimize the sum of α_i and maximize the sum of β_i simultaneously. By adding α_i into (a), we have:

(b) $A_i X \leq b + \alpha_i, A_i \in B$ and $A_i X > b - \alpha_i, A_i \in N$.

However, by considering β_i , we can rewrite (b) as

(c) $A_i X = b + \alpha_i - \beta_i, A_i \in B$ and $A_i X = b - \alpha_i + \beta_i, A_i \in N$.

Our two-criterion linear programming model is stated as

(d) Minimize $\sum_i \alpha_i$ and Maximize $\sum_i \beta_i$
Subject to

$$A_i X = b + \alpha_i - \beta_i, A_i \in B,$$

$$A_i X = b - \alpha_i + \beta_i, A_i \in N,$$

Where A_i is given, X and b are unrestricted, α_i and $\beta_i \geq 0$, see Shi et al. (2001, p. 429).

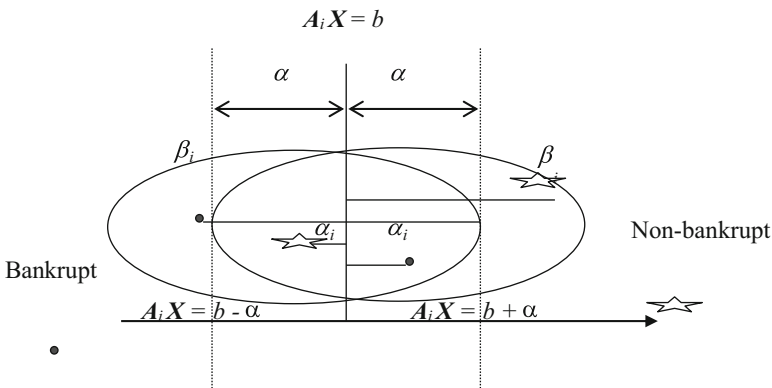
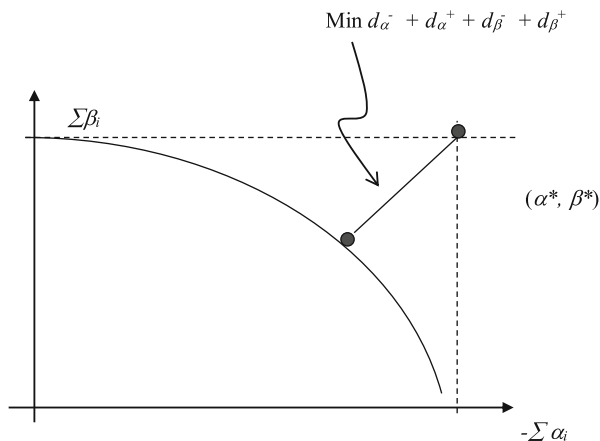


Fig. 1 A Two-class MCLP Model

Fig. 2 Compromise formulation



The above MCLP model can be solved in many different ways. One method is to use the compromise solution approach as Shi and Yu (1989) and Shi (2001) did, to reform model (d) by systematically identifying the best trade-offs between $-\sum_i \alpha_i$ and $\sum_i \beta_i$. To show this graphically, we assume the ideal value of $-\sum_i \alpha_i$ is $\alpha^* > 0$ and the ideal value of $\sum_i \beta_i$ is $\beta^* > 0$. Then, if $-\sum_i \alpha_i > \alpha^*$, we define the regret measure as $-d_{\alpha}^+ = \sum_i \alpha_i + \alpha^*$; otherwise, it is 0. If $-\sum_i \alpha_i < \alpha^*$, the regret measure is defined as $d_{\alpha}^- = \alpha^* + \sum_i \alpha_i$; otherwise, it is 0. Thus, we have (i) $\alpha^* + \sum_i \alpha_i = d_{\alpha}^- - d_{\alpha}^+$, (ii) $|\alpha^* + \sum_i \alpha_i| = d_{\alpha}^- - d_{\alpha}^+$, and (iii) $d_{\alpha}^-, -d_{\alpha}^+ \geq 0$. Similarly, we derive $\beta^* - \sum_i \beta_i = d_{\beta}^- - d_{\beta}^+$, $|\beta^* - \sum_i \beta_i| = d_{\beta}^- + d_{\beta}^+$, and $d_{\beta}^-, d_{\beta}^+ \geq 0$ (see Fig. 2). An MCLP model has been gradually evolved as

(e) Minimize $d_{\alpha}^- + d_{\alpha}^+ + d_{\beta}^- + d_{\beta}^+$
Subject to

$$\alpha^* + \sum_i \alpha_i = d_{\alpha}^- - d_{\alpha}^+,$$

$$\beta^* - \sum_i \beta_i = d_{\beta}^- - d_{\beta}^+,$$

$$A_i X = b + \alpha_i - \beta_i, A_i \in B,$$

$$A_i X = b - \alpha_i + \beta_i, A_i \in N,$$

Where A_i , α^* , and β^* are given, X and b are unrestricted, α_i , β_i , d_{α}^- , d_{α}^+ , d_{β}^- , $d_{\beta}^+ \geq 0$.

The data separation of the MCLP classification method is determined by solving the above problem. Two versions of actual software have been developed to implement the MCLP method on large-scale databases. The first version is based on the well-known commercial SAS platform (SAS/OR 1990). In this software, the SAS

codes, including the SAS linear programming procedure, are used to solve (e) and produce the data separation. The second version of the software is written by C++ language running on a Linux platform (Kou and Shi 2002).

The reason for developing a Linux version of the MCLP classification software is that the majority of database vendors, such as IBM are aggressively moving to Linux-based system development. This Linux version follows the trend of information technology. Because many large companies currently use an SAS system for data analysis, the SAS version is also useful to conduct data mining analysis under the SAS environment.

76.4 Data Collection and Research Design

In this, study we present the MCLP approach for our bankruptcy study (Kwak et al. 2006a) for illustration purpose and later we will report the summary of other countries' studies from our previous studies. We collected bankrupt firms from 1992 to 1998 and an average of 5.84 times more than the number of matching firms by the size and industry to emulate a real-world bankruptcy situation. We collected financial data from Research Insight (RI) and selected companies listed in three major stock exchanges: New York Stock Exchange, American Stock Exchange, and NASDAQ. We checked other sources such as the Wall Street Journal to make sure we include all bankrupt firms around this time.

We include in our sample companies only up to and including 1998 data from 1992 as there were not enough bankrupt companies after 1998 with all the data available. We ran our model every year since the time period is a factor unless our research design controls all macroeconomic factors.

We used Altman's (1968) original variables and Ohlson's (1980) variables for bankruptcy prediction. Altman (1968) used five ratio variables in his study and Ohlson (1980) used nine variables in his study. We can include more macroeconomic variables later without difficulty in our model.

We first ran our MCLP data mining model using five variables for each year as Altman (1968) did in his original study. We wanted to compare the results of our study with the results of Altman's (1968) study to show the prediction accuracy of our study. Then we ran our MCLP data mining model using nine variables as Ohlson (1980) did in his study. Our goal is to show the effectiveness of our model compared with Altman's (1968) and Ohlson's (1980).

Table 1 shows the descriptive statistics that are similar to Altman's (1968) and Ohlson's (1980) previous studies. As you can see in Panel B, all variables between bankrupt and non-bankrupt firms are significantly different except for the size and market value of equity/book value of total debt variables. The size variable is not significantly different since we matched the control firms based on the size and industry. The market value of equity/book value of total debt is marginally significant. However, the size and changes in net income variables are significantly different compared with Ohlson's (1980) original study and our study because of different time periods. Both variables are significantly negative in our study. We used

Table 1 Descriptive statistics

Panel A: Bankrupt firms						
Variables	N	Mean	Std Dev	Min	Max	
Size	91	−33.34	1.48	−37.03	28.90	
TL/TA	91	1.27	1.38	0.11	8.60	
WCA/TA ^a	96	−0.18	0.93	−6.67	0.80	
CL/CA	91	2.45	6.26	0.11	53.11	
NI/TA	96	0.36	0.48	0.00	1.00	
FUTL	91	−0.83	2.62	−23.99	0.32	
INTWO	90	−0.18	0.49	−3.18	0.76	
OENEG	96	0.85	0.35	0.00	1.00	
CHIN	77	−5.33	15.01	−108.13	−0.23	
RE/TA	96	−2.53	8.99	−82.32	0.26	
EBIT/TA	96	−0.43	1.07	−8.09	0.29	
MKV/TD	96	1.52	3.02	0.00	17.89	
SALE/TA	96	1.70	1.12	0.20	6.10	
Panel B: Non-bankrupt firms						
Variables	N	Mean	Std Dev	Min	Max	t-value
Size	528	−33.19	1.42	−38.16	−29.36	0.96
TL/TA	528	0.51	0.70	0.01	9.05	−8.02***
WCA/TA	528	0.24	0.68	−8.10	0.88	5.23***
CL/CA	528	0.84	1.94	0.03	29.04	−4.76***
NI/TA	528	0.05	0.22	0.00	1.00	−10.18***
FUTL	528	−0.11	0.46	−5.23	0.37	5.86***
INTWO	522	0.02	0.82	−7.31	3.08	2.34**
OENEG	528	0.54	0.50	0.00	1.00	−5.84***
CHIN	368	−1.32	3.85	−60.19	0.94	4.48***
RE/TA	528	−0.93	3.30	−36.69	1.08	3.12**
EBIT/TA	528	−0.07	0.44	−5.00	0.63	5.61***
MKV/TD	528	11.67	55.64	0.02	1208.14	1.79*
SALE/TA	528	1.20	0.82	0.00	6.26	−5.23***

^aThis is the same variable in both studies

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.001$

Variable descriptions:

Size = Log (Total Assets/Gross Domestic Products)

TL/TA = Total Liabilities/Total Assets

WCA/TA = Working Capital Divided by Total Assets

CL/CA = Total Current Liabilities/Total Current Assets

OENEG = if $TL/TA > 1$ then $OENEG = 1$; else $OENEG = 0$

NI/TA = Net Income/Total Assets

FUTL = Funds from Operations/Total Liabilities

INTWO = if $Net\ Income < 0$ or $lag(Net\ Income) < 0$ then $INTWO = 1$; else $INTWO = 0$

CHIN = $(Net\ Income - lag(Net\ Income)) / [absolute(Net\ Income) + absolute(lag\ Net\ Income)]$

WCATA = Working Capital/Total Assets

RE/AT = Retained Earnings/Total Assets

EBIT/AT = Earnings Before Interest and Taxes/Total Assets

MKVAL/TL = Market Value of Equity/Book Value of Total Debt

SALE/TA = Sales/Total Assets

the gross domestic product of each year instead of the GNP price-level index as a deflator, with a base value of 100 for 1968 as did the Ohlson (1980) study.

Compared with Altman's (1968) study, the signs of working capital/total assets, retained earnings/total assets, and earnings before interest and taxes variables are the same for the bankrupt firms of our study, but directions are mixed in non-bankrupt firms. Since the time period is quite different, this comparison cannot be meaningful.

Table 2 presents our MCLP model results using Altman's (1968) five variables by each year and overall years. In our study, we tried the MCLP data mining approach using Altman's (1968) five variables with more numbers of control firms since

Table 2 Predictability of the MCLP model using Altman's five variables

One year prior to bankruptcy				
1992	N	Number correct	Percent correct	Percent error
Type-I	21	12	57.14	42.86
Type-II	70	68	97.14	2.86
Overall prediction rate			87.91	12.09
1993	N	Number correct	Percent correct	Percent error
Type-I	15	11	73.33	26.67
Type-II	74	73	98.65	1.35
Overall prediction rate			94.38	5.62
1994	N	Number correct	Percent correct	Percent error
Type-I	12	6	0.50	0.50
Type-II	63	61	96.83	3.17
Overall prediction rate			89.33	10.677
1995	N	Number correct	Percent correct	Percent error
Type-I	8	4	0.50	0.50
Type-II	58	58	100.00	0.00
Overall prediction rate			93.94	6.06
1996	N	Number correct	Percent correct	Percent error
Type-I	15	13	86.67	13.33
Type-II	96	95	98.95	1.04
Overall prediction rate			97.30	2.70
1997	N	Number correct	Percent correct	Percent error
Type-I	12	8	66.67	33.33
Type-II	70	69	97.14	2.86
Overall prediction rate			92.68	7.32
1998	N	Number correct	Percent correct	Percent error
Type-I	9	4	44.44	55.56
Type-II	89	86	96.63	3.37
Overall prediction rate			91.84	8.16
All years	N	Number correct	Percent correct	Percent error
Type-I	91	42	46.15	53.85
Type-II	521	500	95.97	4.03
Overall prediction rate			88.56	11.44

Table 3 Type-I and Type-II errors for re-estimated Ohlson’s Model using 1990s data

Likelihood ratio	Year	Percent correct	Percent error
38.47	1992	92.5%	7.4%
39.77	1993	96.1	3.7
35.03	1994	98.3	1.7
44.10	1995	100.0	0.0
32.34	1996	93.6	6.1
29.41	1997	94.5	5.5
20.07	1998	93.3	6.7
109.04	All years	79.6	19.7

Altman’s (1968) approach of using the same number of control firms is criticized by Ohlson and others for an upward bias of the prediction error rate. For the overall prediction rate, the lowest prediction rate was 87.91% in 1992 by combining the Type I and Type II errors as Altman (1968) reported. The highest prediction rate was 97.30% in 1996. The overall prediction rate for all years was 88.56%. Altman (1968) reported a 95% prediction rate in his original study, but this rate dropped to 78.4% in the 1980s holdout sample according to Begley et al. (1996). From these results, our MCLP model shows that our approach is a lot more efficient and effective in bankruptcy prediction with current financial data. There are some variations by each year in terms of the Type-I error rate. However, Type-II error rates are very consistent and high (average of 95.67%). In the case of bankruptcy prediction, Type-I error is more costly as we discussed before. From the results of Table 2, we can conclude that the correct prediction rate in relation to Type-I error is very time-sensitive, which means bankruptcy prediction 1 year prior to bankruptcy depends on the specific year or other macroeconomic factors. We may need to extend this study using current bankruptcy firm data to verify our findings.

Table 3 reports overall prediction rates of the logit model based on Type-I and Type-II errors using Ohlson’s (1980) nine variables model with an equal number of bankrupt firms and control firms for each year and overall years. The lowest prediction rate was 92.5% in 1992 and the highest rate was 100% in 1995. However, the overall prediction rate for all years was 79.6%, which is similar to other previous studies. Begley et al. (1996) reported similar results when the re-estimated Ohlson (1980) model is applied to the 1980s holdout sample. This table is presented here to compare with the results of our MCLP data mining model.

We applied the MCLP data mining approach using Ohlson’s (1980) nine variables with their original research design of a higher number of control firms and their results are reported in Table 4. We tried Altman’s (1968) five variables model with an equal number of control firms as they did in the original study and their results are reported in Table 4 for comparison purposes.

Table 4 reports Type I and Type II error rates of both models. If we consider that the costs of Type I errors outweigh the costs of Type II errors in a bankruptcy situation, our focus of prediction error rate should be on Type I errors. The prediction rate of Type I errors is increasing and the prediction rate of Type II errors is

Table 4 MCLP data mining model

Year		Altman Model			Ohlson Model		
		N	Number correct	Percent correct (%)	N	Number correct	Percent correct (%)
1992	Type-I	21	15	71	21	11	52
	Type-II	21	17	81	70	66	94
1993	Type-I	15	13	87	15	13	87
	Type-II	15	13	87	74	72	97
1994	Type-I	12	8	67	12	9	75
	Type-II	12	10	83	63	62	98
1995	Type-I	8	8	100	8	6	75
	Type-II	8	8	100	58	57	98
1996	Type-I	15	14	93	15	8	53
	Type-II	15	14	93	96	93	97
1997	Type-I	12	10	83	12	7	58
	Type-II	12	11	92	70	69	99
1998	Type-I	9	7	78	9	4	44
	Type-II	9	8	89	89	87	98
All data	Type-I	126	109	87	91	39	43
	Type-II	91	62	68	521	492	94

decreasing in Altman's (1968) model compared with the model with more numbers of control firms. This is an interesting result that suggests Altman's (1968) original model could be upward biased and he should use more control firms to represent the real-world situation of bankruptcy. However, in the case of Ohlson's (1980) nine variables model, the Type I error prediction rate is decreasing, and the Type II error prediction rate is increasing in our MCLP approach. From the results of the overall prediction rate, Ohlson's (1980) model is similar to Altman's (1968) model (86.76% vs. 88.56%) using more control firms, but our MCLP model using Ohlson's (1980) nine variables is better than the overall prediction rate of the logit model (79.6%). This sustains our findings in this study. From the above results, our MCLP data mining approach performs better than Altman's (1968) model but provides similar or better results to Ohlson's (1980) nine variables model.

Figure 3 presents the cumulative distribution of bankrupt and non-bankrupt firms for all years using Ohlson's (1980) model. The picture is presented to visualize the results of bankruptcy prediction. As you can see from Fig. 3, our MCLP data mining approach is stable about 43% for the Type I prediction rate, but the Type II prediction rate is more than 90% for the overall years.

In our other paper, we used Japanese data using the MCLP data mining approach to predict bankruptcy during 1989–1999 (Kwak et al. 2006a). Although our overall prediction rate using the MCLP approach is somewhat less than that reported by Kwak et al. (2006b), the primary difference between Japanese study and that of Kwak et al. (2006b) is that the difference between Japanese data and US data.

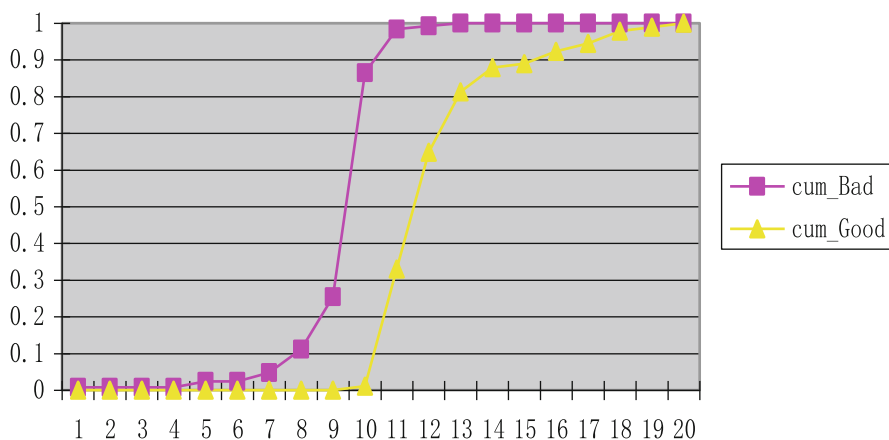


Fig. 3 Cumulative distribution of bankrupt and non-bankrupt firms

Differences in the predication rates may be caused by accounting systems and its information environment of the two countries.

Kwak et al. (2012) also used the MCLP data mining approach to predict bankruptcy using 1997–2003 financial data of Korean bankrupt firms. Overall, the prediction rates in Korean study show 88.7%. Therefore, our MCLP data mining model is very efficient and prediction rates are reasonably good compared to other models.

tWe used other data mining models and the logit model to predict bankruptcy in China (Kwak et al. 2014). Using 1999–2007 Chinese bankruptcy data, our results suggest that Chinese firms with current liquidity problems and firms experiencing a decline in profits are more likely to file for bankruptcy. Overall prediction rates are lower than those of previous USA, Japanese, or Korean bankruptcy studies as shown in our previous studies. China’s capital market has become more dynamic and diverse. We may need to incorporate productivity enhancement and effective strategy implementation for Chinese firms to improve bankruptcy prediction rates (Bryan et al. 2013). In addition, we may use time-series data as suggested by Shumway (2001) and macroeconomic factors to improve our bankruptcy model (Nam et al. 2008).

76.5 Summary and Conclusions

In this chapter, we present empirically the usefulness of Multiple Criteria Linear Programming (MCLP) data mining approaches to predict bankruptcy across countries such as Japan, Korea, and China compared with those of US study. Data mining has proven to be a useful tool in accounting and finance for various prediction and classification studies by several researchers. By using traditional financial ratios and other financial variables such as Altman’s (1968) five variables, using MCLP data mining for bankruptcy studies generally shows better prediction rates than using

traditional logit or probit analysis. However, US data empirically show better prediction rates than Japanese, Korean, and Chinese data around similar time periods. This could be a more reliable information environment in the USA.

Our next research will compare our MCLP data mining approach with other current data mining approaches with current data. We may also include nonfinancial data such as stock market return, qualified audit opinions, internal control weakness data, and other macroeconomic data to improve bankruptcy prediction rates.

References

- Altman, E. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23 (3): 589–609.
- Altman, E., R.G. Haldeman, and P. Narayanan. 1977. ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance* 1 (1): 29–54.
- Amani, F.A., and A.M. Fadlalla. 2017. Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems* 24: 32–58.
- Batena, L., and F. Asghari. 2020. Bankruptcy prediction using logit and genetic algorithm methods: A comparative analysis. *Computational Economics* 55: 335–348.
- Beaver, W. 1966. Financial ratios as predictors of failure: Empirical research in accounting. selected studies. *Supplement to the Journal of Accounting Research*: 71–111.
- Begley, J., J. Ming, and S. Watts. 1996. Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. *Review of Accounting Studies* 1: 267–284.
- Berry, M.J.A., and G. Linoff. 1997. *Data mining techniques for marketing, sales, and customer support*. New York: Wiley Computer Publishing.
- Bryan, D., G.D. Fernando, and A. Tripathy. 2013. Bankruptcy risk, productivity, and firm strategy. *Review of Accounting and Finance* 12 (4): 309–326.
- Chan, C., and B. Lewis. 2002. A basic primer on data mining. *Information Systems Management*: 56–60.
- Freed, N., and F. Glover. 1986. Evaluating alternative linear programming models to solve the two-group discriminant problem. *Decision Sciences* 17: 151–162.
- Grice, J.S., and M.T. Dugan. 2001. The limitations of bankruptcy prediction models: Some cautions for the researcher. *Review of Quantitative Finance and Accounting* 17: 151–166.
- Han, J., and M. Kamber. 2001. *Data mining: Concepts and techniques*. San Francisco: Morgan Kaufmann Publishers.
- Horak, J., J. Vrbka, and F. Suler. 2020. Support vector machine methods and artificial neural networks used for the development of bankruptcy prediction models and their comparison. *Journal of Risk and Financial Management* 13 (60): 1–15.
- Kim, S.B., M. Mun, and S.J. Bae. 2018. Data depth-based support vector machines for predicting corporate bankruptcy. *Applied Intelligence* 48: 791–804.
- Klepac, V., and D. Hampel. 2018. Predicting bankruptcy of manufacturing companies in EU. *Finance* 21: 159–174.
- Korol, T. 2020. Assessment of trajectories of non-bankrupt and bankrupt enterprises. *European Research Studies Journal* 23 (4): 1113–1135.
- Kou, G., and Y. Shi. 2002. *Linux based multiple linear programming classification program: version 1.0*. College of Information Science and Technology, University of Nebraska-Omaha, Omaha.
- Kou, G., X. Liu, Y. Peng, Y. Shi, M. Wise, and W. Xu. 2003. Multiple criteria linear programming approach to data mining: Models, algorithm designs and software development. *Optimization Methods and Software* 18: 453–473.

- Kwak, W., S. Eldridge, S. Yong, and K. Gang. 2006a. Bankruptcy prediction for Japanese firms: Using Multiple Criteria Linear Programming data mining approach. *International Journal of Business Intelligence and Data Mining* 1 (4): 401–416.
- Kwak, W., Y. Shi, and J.J. Cheh. 2006b. Firm bankruptcy prediction using Multiple Criteria Linear Programming data mining Approach. *Advances in Investment Analysis and Portfolio Management* 2: 27–49.
- Kwak, W., S. Eldridge, S. Yong, and K. Gang. 2009. Predicting material weaknesses in internal control systems after the Sarbanes-Oxley Act using Multiple Criteria Linear Programming and other data mining approaches. *Journal of Applied Business Research* 25: 105–118.
- . 2011. Predicting auditor changes using financial distress variables and the Multiple Criteria Linear Programming (MCLP) and other data mining approaches. *Journal of Applied Business Research* 27: 73–84.
- Kwak, W., Y. Shi, and G. Kou. 2012. Bankruptcy prediction for Korean firms after the 1997 financial crisis: Using a multiple criteria linear programming data mining approach. *Review of Quantitative Finance and Accounting* 38 (4): 441–453.
- Kwak, W., X. Cheng, J. Ni, Y. Shi, G. Gong, and N. Yan. 2014. Bankruptcy prediction for Chinese firms: Comparing data mining tools with logit analysis. *Journal of Modern Accounting and Auditing* 10 (10): 1030–1037.
- Mousavi, M.M., and J. Ouenniche. 2018. Multi-criteria ranking of corporate distress prediction models: Empirical evaluation and methodological contributions. *Annals of Operation Research* 271: 853–886.
- Nam, C., T. Kim, N. Park, and H. Lee. 2008. Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting* 27 (6): 493–506.
- Nanda, S., and P. Pendharkar. 2001. Linear models for minimizing misclassification costs in bankruptcy prediction. *International Journal of Intelligent Systems in Accounting, Finance & Management* 10: 155–168.
- Ohlson, J. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18 (1): 109–131.
- Platt, D.H., and M.B. Platt. 1990. Development of a class of stable predictive variables: The case of bankruptcy prediction. *Journal of Business Finance & Accounting Spring*: 31–51.
- Sam, R., J. Osleeb, and K. Si. 2016. The maximal cover location model with hedging: Siting facilities under uncertainty, a lead poisoning screening network for the Dominican Republic. *International Regional Science Review* 39: 77.
- SAS/OR User's Guide. 1990. SAS Institute Inc., Cary, NC.
- Segal, S.Y. 2016. Accounting frauds – review of advanced technologies to detect and prevent frauds. *Economics and Business Review* 2 (4): 45–64.
- Shi, Y. 2001. *Multiple criteria multiple constraint-levels linear programming: Concepts, techniques and applications*. River Edge: World Scientific Publishing.
- Shi, Y., M. Wise, M. Luo, and Y. Lin. 2001. Data mining in credit card portfolio management: A multiple criteria decision making approach. In M. Koksakan and S. Zionts (Eds.), *Multiple Criteria Decision Making in the New Millennium*, 427–436.
- Shi, Y., and P.L. Yu. 1989. Goal setting and compromise solutions. In *Multiple criteria decision making and risk analysis using microcomputers*, ed. B. Karpak and S. Zionts, 165–204. Berlin: Springer-Verlag.
- Shi, Y., Y. Peng, W. Xu, and X. Tang. 2002. Data mining via multiple criteria linear programming: Applications in credit card portfolio management. *International Journal of Information Technology & Decision Making* 1 (1): 131–151.
- Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business* 74 (1): 101–124.
- Witten, I.H., and E. Frank. 2000. *Data Mining: Practical machine learning tools and techniques with java implementations*. San Francisco: Morgan Kaufmann Publishers.



Application of Difference-in-Differences Strategies in Finance: The Case of Natural Disasters and Bank Responses

77

James R. Barth, Kang Bok Lee, Xuan Shen, and Yeo Song Yoon

Abstract

Natural disasters are negative shocks that can severely disrupt the communities in which they occur. Disasters like hurricanes, tornados, floods, wildfires, and earthquakes, moreover, can cause severe property damage, including damages to homes, businesses, and automobiles. An important issue that arises is whether branches of banks in communities affected by natural disasters raise deposit rates to attract additional deposits in response to any deposit withdrawals and to meet any increase in loan demand for the rebuilding that takes place. Studies of the responses of banks to natural disasters increasingly find it useful to rely on a difference-in-differences (DID) identification strategy. The contribution of our entry is to examine how different choices that can be made affect the empirical results. Importantly, as our empirical results indicate, the discretion that a researcher uses in terms of the choices made at different stages do indeed produce different conclusions about the impact of natural disasters on bank deposit rates. If we do not match branches of branches affected by natural disasters with those

J. R. Barth (✉)

Auburn University and Milken Institute, Auburn, AL, USA

Department of Finance, Harbert College of Business, Auburn University, Auburn, AL, USA

e-mail: barthjr@auburn.edu

K. B. Lee

Department of Systems and Technology, Harbert College of Business, Auburn University, Auburn, AL, USA

e-mail: kbl0009@auburn.edu

X. Shen

Quantitative Risk Analytics, Regions Bank, Birmingham, AL, USA

e-mail: xuan.shen@regions.com

Y. S. Yoon

Department of Finance, Harbert College of Business, Auburn University, Auburn, AL, USA

e-mail: yzy0087@auburn.edu

in adjacent communities not affected the results indicate that natural disasters have a statistically negative effect on deposit rates without matching and with a low degree of matching. However, when we use a medium or a high degree of matching there is no statistically significant effect. Moreover, when we use two different matching methods, the results differ. In the case of PSM, we find a statistically significant effect, but no effect in the case of CEM.

Keywords

Natural Disasters · Banks · Deposit Rates · Difference-in-Differences · Matching · Bank Branches · Common Trend

JEL Classification

G20 · G21 · C10 · C40

77.1 Introduction

Natural disasters are negative shocks that can severely disrupt the communities in which they occur. Such disasters, moreover, are not rare events. Indeed, since the 1990s, there has been a clear acceleration in disasters, averaging more than 300 county-level events per year over the period 2000–2013. Furthermore, disasters like hurricanes, tornados, floods, wildfires, and earthquakes can cause severe property damage, including damages to homes, businesses, and automobiles. Although insurance and government support cover some of the reconstruction and replacement costs, individuals and business owners suffering damages may withdraw deposits at banks and apply for loans to help cover all the costs.

An interesting and important issue that arises is the impact of natural disasters on banks. For example, do individuals and business owners withdraw deposits in response to such disasters to obtain funds to cover expenses associated with recovery efforts? And do branches of banks in communities affected by natural disasters raise deposit rates to attract additional deposits in response to any deposit withdrawals and to meet any increase in loan demand for the rebuilding that takes place? Studies of the responses of banks to natural disasters, and other credibly exogenous sources of variation, increasingly find it useful to rely on a difference-in-differences (DID) identification strategy. The reason is that such a strategy provides a stronger foundation to achieve the identification of causal relationships rather than simply correlations.¹ However, when using such a strategy one assumes there is a good match between the treatment group (e.g., bank branches in communities affected by disasters) and the control group (e.g., bank branches in communities not affected by them). The purpose of a good match is to reduce potential confounding by improving the comparability of units (e.g., bank branches) in the treatment and

¹ Angrist and Pischke (2010, p. 14) state that DID is “... probably the most widely applicable design-based estimator.”

control groups. One also assumes that there is a common trend (e.g., in deposit rates) for the two groups before the treatment.

The contribution of our entry is to analyze the effect of natural disasters on the deposit rates of banks. However, for a nonexperimental method, like DID that we will be using, there are several stages in the research process where a researcher may exercise discretion (Brodeur et al. 2020). We, therefore, examine how different choices that can be made affect the empirical results. This is important due to the concern that a researcher might engage in *p*-hacking, which refers to the practice used to generate “better” *p*-values, perhaps, though not necessarily, in response to a publication bias or the difficulty of publishing statistically insignificant results (Brodeur et al. 2020).² Specifically, we show there are choices regarding the matching of branches and the type of matching method employed when using DID. Thus, we compare the results from estimating a basic model of deposit rate responses to natural disasters using a DID identification strategy based on several different choices made. The choices we make are with and without matching of bank branches as well as using “poor” and “good” matches, adjectives that are defined later. We also choose different matching methods to determine whether they produce the same or different results. Lastly, we choose a sample with a pre-disaster common trend in deposit rates in counties experiencing natural disasters and those in adjacent counties not experiencing them as well as a subsample with no pre-disaster common trend. Importantly, as our empirical results indicate, the discretion that a researcher uses in terms of the choices made at different stages do indeed produce different conclusions about the impact of natural disasters on bank deposit rates. Indeed, some choices could be considered as *p*-hacking. Yet, the most reliable results indicate that natural disasters have no statistically significant effect on bank deposit rates. This finding might produce publication bias, whereby the lack of statistical significance lowers the probability of publication (Brodeur et al. 2020).

The remainder of the entry proceeds as follows. In the next section, we describe the motivation for our study. The third section explains the data, and its various sources, used in the empirical work. This is followed in the fourth section with a discussion of matching methods and the quality of the methods using an imbalance measure. The fifth section explains our research design. The basic model to be estimated as well as the empirical results are presented and discussed in the sixth section. The seventh and last section contains the conclusions.

77.2 Motivation

Our study of the effect of natural disasters on bank deposit rates is motivated in part by a recent paper by Dlugosz et al. (2020). They use a DID identification strategy to examine whether branches of banks in communities affected by natural disasters can set deposit rates locally and thereby be in a position to raise those rates to attract additional deposits to meet an increase in loan demand for the reconstruction and replacement efforts that

²Brodeur et al. (2020, p. 3658), moreover, state that “. . . papers using IV and DID are identified as particularly problematic with respect to *p*-hacking and publication bias problems.”

take place after such disasters. They find that branches that set rates locally do increase deposit rates more in counties affected by natural disasters than branches in adjacent counties not experiencing them. We adopt a similar approach to analyzing the response of deposit rates at branches in counties suffering natural disasters, including their model specification, data sources, and sample period. However, they chose to implement a triple DID strategy and rely on a single matching method, while we choose to employ a DID strategy and rely on two matching methods.

Also, our study, in contrast to Dlugosz et al. (2020), is partly motivated by the discretion a researcher can exercise at different stages in the implementation of the DID identification strategy (Brodeur et al. 2020). We therefore compare and contrast the results from the different choices that one can make, including choosing with and without matching, and poor and better matching, of bank branches in counties with natural disasters and those in adjacent counties without them. Also, we consider the effect on the results of matching banks when one can choose between two different matching methods. Lastly, we examine whether there is a difference in results for rates when one can choose either with or without a common trend for the two groups of branches before the occurrence of natural disasters. Importantly, we find that the empirical results do indeed differ depending upon the research design when using DID to determine whether bank deposit rates are different for banks in counties experiencing natural disasters and those in adjacent counties not experiencing them. Our findings have important implications for the degree to which one can rely upon different DID strategies for not only a study like that of Dlugosz et al. (2020), but also other studies using DID techniques to access the response of a bank, or another firm, to various types of exogenous shocks.

77.3 Data

The data for our study combines several datasets covering the period 1999 to 2014, which is the same period used by Dlugosz et al. (2020). These include information on natural disasters, deposit rates for each bank branch, bank balance-sheet variables, and branch location. We now describe in more detail the construction of the data used in the empirical work.

77.3.1 Natural Disasters

The data on natural disasters are from the Federal Emergency Management Agency (FEMA).³ FEMA provides the declaration date of the disaster, location (county and state), and type of disaster. We include the following types of disasters in the analysis: fires, floods, hurricanes, snow, severe storms, tornados, and other disasters. We exclude disasters in territories such as Puerto Rico. The total number of disasters from 1999 to 2013 is 4,287, ranging from a low of 60 in 1999 to a high of 482 in

³See <https://www.fema.gov/data-sets>

2007. This translates into 24,688 natural disasters in counties for our sample period. This number is reduced to 14,751 disasters in counties once we remove counties with multiple disasters in the same month and those suffering any other disaster in a 7-month window other than but including the disaster month (i.e., 3 months before and 3 months after a disaster as well as the month of the disaster). For analyzing the impact of disasters, we form a cohort including a disaster-affected county and its adjacent non-affected counties to compare the deposit-rate responses of the branches within each of the counties. Our treatment group consists of the branches within a disaster-affected county and the control group consists of the branches within its adjacent non-affected counties. We keep those cohorts for which there are at least one affected and one unaffected county. This further reduces the sample to 7,894 counties with natural disasters and adjacent counties without disasters.

77.3.2 Deposit Rates

We obtain deposit rate data from RateWatch, a firm that provides weekly branch-level data for such rates. Information is available for the location of branches (county and state), unique branch IDs assigned by the Federal Deposit Insurance Corporation (FDIC), and type of deposit products. We use the deposit rates of each branch for 12-month certificates of deposit requiring a minimum balance of \$10,000 (CDs). Since the deposit rate data are weekly, we use the monthly average deposit rate for each branch in our analysis.

77.3.3 Deposit Data

Deposit data at the branch-level are available from the Summary of Deposits (SOD), FDIC. SOD provides an annual survey of branch deposits as of June 30, from 1999 to 2013. The data also includes geographic information (county and state) of branches and bank headquarters as well as unique branch IDs assigned by the FDIC. We keep the cohorts within which at least one affected and at least one unaffected county exists and drop those branches if for any month within the event window they are missing.

77.3.4 Bank Variables

The data for the bank variables are available from the quarterly Call Reports and the Statistics on Depository Institutions (SDI), FDIC. The following three bank control variables, also used by Dlugosz et al. (2020), are used in our estimation: (1) 1 if a bank is a community bank (*Small*),⁴ 0 otherwise, (2) 1 if a bank has more than 65%

⁴ Actually, Dlugosz et al. (2020) define *Small* to be a dummy variable that equals 1 if a bank has less than \$2 billion in assets

of its deposits from a single county (*Local*), 0 otherwise, and (3) 1 if a county is in the top quartile of deposits among the counties in which a bank has branches (*Important Market*), 0 otherwise. The values of the control variables are determined at the end of the quarter before the month of the natural disaster. The bank control variables are transformed into monthly values by interpolating their quarterly values.

With the unique branch IDs and geographic information, we can match and merge SOD data with both RateWatch deposit rates and FEMA natural disaster data at the county level. In the next section, we discuss the matching techniques and imbalance measure used in our empirical work.

77.4 Matching Methods and Imbalance Measure

Matching Methods. Random assignment is required to construct a valid causal inference, as it provides a perfect balance in covariates between treatment and control groups. With observational datasets, obtaining valid causal inference is a challenge due to the difference between the treatment and control groups (Heckman et al. 1998). Thus, in the absence of random assignment, the goal of matching is to reduce the observable differences, or, the imbalance between the two groups (Stuart 2010; Iacus et al. 2012). A better balance is achieved when there is a similar distribution of covariates between the two groups. Previous research has shown that large deviations in the distribution of covariates between groups can lead to a bias in estimating the treatment effect (Cochran and Rubin 1973; Rubin 1973; Rubin and Thomas 2000). A matching method can be used to obtain a well-matched sample that reduces statistical bias in estimating the treatment effect and thereby improving the estimation of causal effects (Ho et al. 2007; Stuart 2010; Iacus et al. 2012). In this regard, we employ two matching methods: Coarsened Exact Matching (CEM) and Propensity Score Matching (PSM) to estimate the DID parameter after matching the branches. The first method, CEM, is one of a new generalized class of matching methods called the monotonic imbalance bounding (MIB) method. This involves preprocessing data by coarsening variables, implementing one-to-one exact matching, and reducing multivariate imbalance measures (Blackwell et al. 2009; Iacus et al. 2011, 2012). The second method, PSM, is one of the most widely used matching methods for causal inference and for analyzing the causal effect of treatments specifically. PSM⁵ was introduced by Rosenbaum and Rubin (1983) and has been implemented in numerous empirical research studies in statistics, economics, finance, and other fields. By implementing parametric models, such as logit or probit models based on the matching variables, the propensity score, which is the likelihood of receiving treatment, is estimated to compute the causal effect.

Imbalance Measure. When using the two matching techniques, one can determine the quality (i.e., good versus poor) of the match of variables for the two groups by calculating a multivariate imbalance measure. Specifically, we can determine which covariate balance based on different matches of variables between the treatment (i.e.,

⁵PSM is a member of the class of matching methods called equal percent bias reducing (Rubin and Thomas 1992; Rubin and Stuart 2006).

branches in counties suffering natural disasters) and control (i.e., branches in counties in adjacent counties not suffering them) groups is most appropriate for the DID estimation. To gauge the difference in the multivariate distributions between the treatment and control groups, one calculates the multivariate imbalance measure ($\mathcal{L}_1$) (Iacus et al. 2012). $\mathcal{L}_1$ falls between 0 and 1; $\mathcal{L}_1 = 0$ if the distributions perfectly overlap, and $\mathcal{L}_1 = 1$ if the distributions are completely separated. For example, if $\mathcal{L}_1 = 0.4$, then 60% of the two groups overlap. Thus, the lower the imbalance measure, $\mathcal{L}_1$, the better the match.⁶ In our study, a low degree of matching of bank branches is defined as a low reduction (i.e., a decrease of less than 0.02, no reduction, or even an increase) in the multivariate imbalance measure after matching when the measure after matching remains higher than 0.65. In contrast, a high degree of matching is defined as a high reduction (i.e., a decrease of more than 0.6) in the multivariate imbalance measure after matching when the measure after matching is lower than 0.20. For example, when the imbalance measures based on before matching (using the full dataset) and after matching are 0.9 and 0.2, respectively, we classify the case as a high reduction (0.7 is greater than 0.6 and the decrease is more than 0.6). The degree of matching is considered medium when the imbalance measure and the magnitude of reduction after matching remain between the values of the low and high degrees of matching.

Common Trend Assumption. In the empirical DID literature, the common trend assumption is a key identifying assumption because the DID estimator requires that, in the absence of treatment, the differences in the average outcomes between the treatment and the control groups would be constant in the post and pretreatment periods (Angrist and Pischke 2008; Lechner 2011; Mora and Reggio 2019⁷).⁸ Given the purpose of our study, we estimate parameters in our models in which the pre-disaster common trend is satisfied as well as not satisfied for the matched branches. Based on the work of Mora and Reggio (2015), we test the common trend assumption using the `didq` command in STATA.⁹ This test enables us to determine whether our samples satisfy the common trend assumption for using the DID model when several pretreatment periods exist.

77.5 Research Design

Given one of the main purposes of this study, we estimate several variations of a basic DID model. More specifically, we estimate DID models for samples both with and without matching bank branches in counties experiencing natural disasters as well as in adjacent counties not experiencing natural disasters. We also

⁶For more details, see [Appendix 1](#).

⁷Mora and Reggio (2015, p. 1) state that “When there is only one pretreatment period, this assumption is the parallel paths assumption. When there are several pretreatment periods, the assumption is usually referred to as common trends.”

⁸For more details, see [Appendix 2](#).

⁹When using this command, if the *p-value* is over some threshold, such as 0.05, then one cannot reject the null hypothesis that there are common trends.

estimate DID models based on different sets of matching variables for the branches in the two types of counties. Lastly, we estimate DID models in which the pre-disaster common trend is satisfied as well as not satisfied for the matched branches in a similar manner.

Tables 1 and 2 show the basic approach to our empirical estimation. In Table 1, there are six question marks, indicating our objective to determine whether common trends and matching affect the impact of natural disasters on bank deposit rates. We also examine whether there is a difference in the results (Δ Imbalance Measure¹⁰) for low, medium, and high degrees of matching, based on the imbalance measure as described in the previous section. This type of exercise is not typically performed in empirical studies, which raises the question of whether the use of DID is dependent upon the underlying matching of treatment and control groups used and a common trend assumption commonly made to produce reliable results.

Table 2 is quite similar to Table 1. The only difference is that, in the former table, two different matching methods (PSM and CEM) are employed when estimating the models.

The next section discusses the empirical DID models used in estimating the treatment effect and the empirical results.

Table 1 Do degree of matching matter and a common trend matter?

LOW generalizability					
Δ Imbalance measure		Matching methods			
			CEM		
			Set (1)	Set (2)	Set (3)
		No matching	Low	Medium	High
Common trend	Not satisfied	?	?	?	?
	Satisfied	?	?	?	?

Table 2 Do type of matching method and a common trend matter?

HIGH generalizability				
Δ Imbalance measure		Matching methods		
			PSM	CEM
			Set (3)	Set (3)
		No matching	Low	High
Common trend	Not satisfied	?	?	?
	Satisfied	?	?	?

¹⁰ Δ Imbalance Measure = before matching imbalance measure – after matching imbalance measure.

77.6 Models and Empirical Results

77.6.1 Base Scenario – DID Model Without a Matching Method

We use a DID model to estimate the average treatment effect without any matching method. To examine the impact of natural disasters on branch deposit rates, we focus on a 7-month event window (i.e., 3 months before and 3 months after a disaster) centered on the month in which a natural disaster occurred. This is a period that encompasses the occurrence of an external shock (Dlugosz et al. 2020). Given this situation, we postulate the following hypothesis:

Hypothesis: *When a natural disaster occurs, branches within disaster-affected counties are more likely to experience an increase in deposit rates than branches within unaffected adjacent counties.*

To test this hypothesis, we use the following basic regression model:

$$\text{Deposit rate}_{ijct} = \beta_0 + \delta_0 \text{Post}_t + \beta_1 \text{Treated}_{ijc} + \delta_1 \text{Post}_t \times \text{Treated}_{ijc} + \mathbf{X}'_{bj} \mathbf{\Lambda} + e_c + \varepsilon_{ijct} \quad (1)$$

where i, j, c, b , and t denote a branch, a county, a cohort, a bank that owns branch i , and time ($t = 0, 1$), respectively. A cohort, c , is formed when a disaster-affected county is combined with its adjacent non-affected counties. The cohort, c , is added to compare the deposit responses of affected and non-affected branches within a cohort. The $\text{Deposit rate}_{ijct}$ denotes the average deposit rate for branch i within county j in cohort c at time t , where the rate is defined as the average deposit rate for the 4 months after the disaster (including the disaster month; $t = 1$) and the 3 months before the disaster ($t = 0$); Post_t equals 1 for the 4 months after the disaster ($t = 1$) and 0 for the 3 months before the disaster ($t = 0$) (i.e., $\text{event month} = -3, -2, -1, 0, +1, +2, +3$ where $\text{disaster} = 0$. $\text{Post}_t = 1$ if $\text{event month} \geq \text{disaster month}$); Treated_{ijc} equals 1 if branch i is within a disaster-affected county in cohort c ; $\mathbf{X}'_{bj} \mathbf{\Lambda}$ is a vector of bank control variables, as noted earlier; e_c indicates cohort fixed effects; and ε_{ijct} is an idiosyncratic error term. Standard errors are clustered (Cluster SE) at the bank level. In the following sections, subscripts are often omitted for simplicity when their omission does not create any confusion.

Following Bertrand et al.'s (2004) suggestion to mitigate serial correlation, the dataset is aggregated into one observation for the pretreatment period and one for the posttreatment period by taking the average value of all the variables for each branch in a cohort (i.e., $\text{Deposit rate}_{ijct}$). Our control variables include a small bank dummy (*Small*), a local bank dummy (*Local*), and an important market dummy (*Important Market*), representing factors that might affect branch deposit behavior (Dlugosz et al. 2020). The coefficient of interest, δ_1 , denotes the DID estimator and measures the effect of a natural disaster on the deposit rate, which may be positive, negative, or zero.

Table 3 Summary statistics

	<i>N</i>	Mean	SD	25th	Median	75th
<i>Panel A. Full dataset (common trend assumption satisfied)</i>						
12-month CD rate, %	85,948	2.08	1.44	0.90	1.74	3.20
Treated (dummy)	85,948	0.30	0.46	0.00	0.00	1.00
Ln (assets)	85,948	12.99	2.38	11.42	12.31	13.70
Total assets, \$ billions	85,948	29	158	0.09	0.22	0.90
Important market (dummy)	85,948	0.12	0.32	0.00	0.00	0.00
Small (dummy)	85,948	0.77	0.42	1.00	1.00	1.00
Local (dummy)	85,948	0.61	0.49	0.00	1.00	1.00
<i>Panel B. Subset sample (common trend assumption unsatisfied)</i>						
12-month CD rate, %	35,766	2.10	1.43	0.93	1.76	3.18
Treated (dummy)	35,766	0.30	0.46	0.00	0.00	1.00
Ln (assets)	35,766	12.89	2.34	11.34	12.28	13.61
Total assets, \$ billions	35,766	26	148	0.08	0.22	0.82
Important market (dummy)	35,766	0.16	0.37	0.00	0.00	0.00
Small (dummy)	35,766	0.78	0.42	1.00	1.00	1.00
Local (dummy)	35,766	0.63	0.48	0.00	1.00	1.00

To examine whether the empirical results differ with and without a common trend for the two sets of branches, we construct a subset¹¹ of observations in which the pretreatment common trend is not satisfied. In other words, we determine whether our coefficient of interest is affected by the common trend assumption. By testing the common trend assumption (Mora and Reggio 2015), our subset sample is formed with the branch cohorts that are affected by more than three disasters between 1999 and 2013. Using this particular sample, we also estimate the same specification as in Eq. (1) to determine whether the results change.

Table 3 provides a summary of the statistics of our full dataset and subset samples. As can be seen, apart from the difference in sample size, the values of the variables themselves are quite similar in magnitude when measured in the different ways indicated.

77.6.2 Imbalance Measure, Common Trend Assumption, and DID Estimator

To estimate the causal effect of the treatment, we employ CEM as our first matching method and then use the matched branches to estimate the same specification given by Eq. (1). As mentioned earlier, CEM is part of a new generalized class of matching methods called the MIB method, which involves preprocessing data by coarsening

¹¹This subset is obtained from our full sample used to estimate Eq. (1). We test the common trend assumption before aggregating the data and find that the full dataset satisfies the common trend assumption.

variables, implementing one-to-one exact matching, and reducing multivariate imbalance measures (Blackwell et al. 2009; Iacus et al. 2011, 2012). This is the reason we employ CEM, first because this method has desirable statistical properties and second it shows the improvement of the multivariate imbalance measure compare to other existing matching methods (Iacus et al. 2012). We will discuss PSM in Section C.

CEM is performed *before* the average of observations in the 4 months after a disaster and the 3 months before a disaster (i.e., 7 event months) is calculated. We match the same event month in the treatment and control groups (e.g., we match event month 2 in the treatment group with event month 2 in the control group). Performing CEM involves three steps: (1) select the matching variables and calculate the imbalance measure, (2) coarsen variables using coarsening methods (automated or by user), and (3) ensure the imbalance measure is reduced after CEM is performed to find the set of variables yielding best-matched branches. Our matching strategy is to apply three matching variable sets and show how multivariate imbalance measures are reduced after matching using CEM. The three matching variable sets consist of (1) poor or low degree matching variables, (2) poor matching variables plus good or medium degree matching variables A, and (3) poor matching variables plus good matching variables A plus good or high degree matching variables B, as shown in Table 4.

Table 4 Matching variables sets

Δ Imbalance measure		
Low	Medium	High
Set 1	Set 2	Set 3
Poor matching variables	Poor matching variables + good matching variables A	Poor matching variables + good matching variables A good matching variables B
Ln (total assets)	Ln (total assets)	Ln (total assets)
Ln (total deposits)	Ln (total deposits)	Ln (total deposits)
Equity-asset ratio	Equity-asset ratio	Equity-asset ratio
Security-asset ratio	Security-asset ratio	Security-asset ratio
	Loan-asset ratio	Loan-asset ratio
	Residential mortgage-loan ratio	Residential mortgage-loan ratio
	Year ^a	Consumer – loan ratio
	Month ^a	Year ^a
		Month ^a
		Small
		Branch ratio ^b

^aYear and month in the event window

^bFraction of branch of the bank within the county

Poor matching variables (low degree of matching) refer to matching variables that minimally reduce the imbalance measures. Good matching variables (high degree of matching) refer to matching variables that substantially reduce imbalance measures. As shown in Sets 1, 2, and 3, the matching procedure increasingly includes more variables, and changes in the imbalance measures moving from one set to the next are calculated. The three-match samples are then used with the same specification to estimate the parameters for the variables in Eq. (1). By adding good matching variables A and B to produce Set 3, we obtain the biggest reduction in the imbalance measure of the three sets. This enables us to determine how the results from the different estimations are affected by different imbalance measures. Also, by matching the same three matching sets for the subset sample when the pretreatment common trend is not satisfied, we can determine whether the common trend assumption matters for the effect of natural disasters on deposit rates. Accordingly, the first research question to be answered is as follows:

Research Question 1: *Do different sets of matching variables and whether a common trend assumption is satisfied change the estimation results?*

To answer this question, Table 5 presents changes in the imbalance measures (i.e., the imbalance measures from the before matching minus the imbalance measures from after matching) for three sets of matching variables of the two groups of branches for each of the 7 months in the event window, as described earlier, but with a common trend for the treatment and control groups before treatment. As can be seen, there is an increasingly larger improvement in the imbalance measures when more good matching variables are included. Indeed, Set 3 shows a far better match than either of the other two sets of variables when all good matching variables are included.

Table 6 presents the same information as Table 5, but the treatment and control groups do not share a common trend before treatment. Once again, there is an

Table 5 Imbalance measure changes for different matching variables with a common trend

Months before, during, and after disaster	Poor matching variables	Poor matching variables + good matching variables A	Poor matching variables + good matching variables A + good matching variables B
	<i>Set 1</i>	<i>Set 2</i>	<i>Set 3</i>
	Δ Imbalance measure	Δ Imbalance measure	Δ Imbalance measure
−3	−0.009	0.379	0.761
−2	0.004	0.400	0.798
−1	−0.005	0.413	0.804
0	−0.005	0.419	0.802
1	0.010	0.407	0.797
2	−0.003	0.592	0.753
3	−0.011	0.588	0.781

Table 6 Imbalance changes for different matching variables without a common trend

	Poor matching variables	Poor matching variables + good matching variables A	Poor matching variables + good matching variables A + good matching variables B
	<i>Set 1</i>	<i>Set 2</i>	<i>Set 3</i>
Months before, during, and after disaster	Δ Imbalance measure	Δ Imbalance measure	Δ Imbalance measure
-3	-0.019	0.413	0.775
-2	0.011	0.402	0.801
-1	-0.013	0.402	0.832
0	0.016	0.402	0.830
1	-0.022	0.429	0.853
2	0.001	0.636	0.868
3	0.002	0.605	0.892

increasingly larger improvement in the imbalance measures when more good matching variables are included. Set 3 provides the best match among the three sets due to the largest reduction in the imbalance measure.

Given the difference in the goodness of matching the two groups of branches with and without a common trend, does the degree of matching and the presence of a common trend matter for deposit rates? The answer is provided in Table 7. Panel A shows the results with a common trend, while Panel B shows the results without a common trend. As can be seen in Panel A, branches in counties struck by natural disasters experience a significant negative impact on deposit rates as compared to branches in unaffected adjacent counties without matching and with poor matching variables (Set 1). However, when good and all good matching variables are included (Sets 2 and 3), the results are quite different. In this case, there is no significant relationship between deposit rates and natural disasters. In other words, matching matters, but only when it occurs when there is a substantial improvement in the imbalance measure, as shown in 5.

The analysis is repeated with the two groups of branches that do not have a common trend before the occurrence of natural disasters. The results reported in Panel B are essentially the same as those reported in Panel A. In this case, however, the presence of a pretreatment common trend in the two groups of branches does not matter for the empirical results.

Table 8 summarizes our findings for the three different matching sets, with and without pretreatment a common trend. Importantly, the results for no matching and poor or low matching, and good or medium and high matching, are quite different. However, the lack of a common trend in the treatment and control groups does not affect the empirical results. In short, matching and the degree of matching based on the imbalance measure matter, but the violation of a common trend assumption does not matter for our sample of branches and study period.

Table 7 Difference-in-differences results. *Post* = 1 for 4 months after a disaster, including the disaster month (*event month* = +3, +2, +1, 0); *Post* = 0 for 3 months before the disaster (*event month* = −1, −2, −3). The deposit rate is the average of the 4 months after the disaster and the average of the 3 months before the disaster.

Dependent variable:		With matching method		
		CEM		
Deposit rate	Without matching method	Set 1	Set 2	Set 3
<i>Panel A. With a common trend</i>				
Treated $\square \times$ Post	−0.0106**	−0.0092**	−0.0037	−0.0029
Controls	Y	Y	Y	Y
Fixed effects	Y	Y	Y	Y
Cluster SE	Bank	Bank	Bank	Bank
Number of observations	85,948	85,235	29,809	10,891
Number of banks	6,471	6,456	4,099	1,148
R-squared	0.923	0.924	0.933	0.953
<i>Panel B. Without a common trend</i>				
Treated $\square \times$ Post	−0.0338***	−0.0269***	−0.0086	−0.0118
Controls	Y	Y	Y	Y
Fixed effects	Y	Y	Y	Y
Cluster SE	Bank	Bank	Bank	Bank
Number of observations	35,766	35,178	9,498	3,132
Number of banks	3,501	3,479	1,706	457
R-squared	0.916	0.915	0.934	0.954

Note: * denotes significance at 10%, ** at 5%, and *** at 1%

Table 8 Summary of results

<i>LOW generalizability</i>					
Δ Imbalance measure		Matching methods			
		No matching	CEM		
			Set 1	Set 2	Set 3
Common trend	Not satisfied	Significant	Significant	Not significant	Not significant
	Satisfied	Significant	Significant	Not significant	Not significant

77.6.3 Different Matching Methods (PSM and CEM), Common Trend Assumption, and DID Estimator

In this section, we investigate our second research question:

Research Question 2: *Do different matching methods and whether a common trend assumption is satisfied change the estimation results?*

To address this question, we employ another matching technique, PSM, which is one of the most widely used matching methods for causal inferences, to analyze the

causal effect of disasters on bank deposit rates. PSM was introduced by Rosenbaum and Rubin (1983) and has been implemented in numerous empirical studies in finance, among other fields of study. Using estimated propensity scores calculated from observed covariates, one can stratify the propensity scores and then use a matching method (e.g., nearest neighbor or radius matching) to calculate the causal effect (Li 2013). However, PSM has an important limitation: it provides a misleading conclusion regarding causal inference. The reason for this is that a true propensity score obtained from the perfect overlap of distributions of the pretreatment variables for the treatment and control groups cannot be identified (Imai et al. 2008; Imai and Ratkovic 2014). Yet, a perfect overlap can be obtained from a true propensity score. This is called the propensity score tautology. Since we implement the estimated propensity score on matching (Diamond and Sekhon 2013), it may not be possible to have a perfect overlap that would enable one to estimate the causal effect. In contrast, Iacus et al. (2012) explain that CEM has more favorable statistical properties: it outperforms other existing matching methods in reducing imbalance, model dependency, estimation error, and bias. Indeed, CEM is a nonparametric model that considerably reduces potential bias and model dependency in comparison to parametric models, such as PSM (Ho et al. 2007; King and Nielsen 2019). King and Nielsen (2019) contend that PSM increases imbalance, inefficiency, and statistical bias.

Once again, we begin our analysis by showing the improvement in imbalance measures for the two matching methods. PSM is performed in the same way as CEM was performed earlier, but the comparison is now based only on Set 3, the set with the highest degree of matching. We apply the nearest neighbor matching using a caliper width of 0.001. Table 9 clearly shows that the reduction in the imbalance measure when matching with PSM is negligible. In contrast, CEM shows a substantial decrease in the imbalance measure. Thus, CEM allows us to obtain a far better match between branches affected by natural disasters and branches in unaffected adjacent counties when a common trend assumption is satisfied. There is essentially no reduction in the imbalance measure for PSM, whereas there is a substantial reduction in the measure for CEM.

Table 9 Imbalance measures for different matching methods with a common trend

Months before, during, and after disaster	PSM	CEM
	Set 3	Set 3
	Δ Imbalance measure	Δ Imbalance measure
−3	0.013	0.761
−2	0.012	0.798
−1	0.011	0.804
0	0.012	0.802
1	0.012	0.797
2	0.012	0.753
3	0.012	0.781

The same comparison of imbalance measures is performed without a common trend assumption satisfied. The same basic results are presented in Table 10 as in Table 9. CEM provides a far better match for the two groups of branches than PSM in terms of imbalance measure reduction. Indeed, there is hardly any reduction in the case of PSM, while there is a substantial reduction when CEM is used as the matching method.

Given that there is a difference in the goodness of the matching between the two groups of branches with and without a common trend, does the type of matching method and whether there is a common trend matter for the effect of natural disasters on deposit rates? The answer is provided in Table 11. Panel A shows the results with

Table 10 Imbalance measures for different matching methods without a common trend

	PSM	CEM
	Set 3	Set 3
Months before, during, and after disaster	Δ Imbalance measure	Δ Imbalance measure
−3	0.015	0.775
−2	0.015	0.801
−1	0.017	0.832
0	0.016	0.830
1	0.016	0.853
2	0.014	0.868
3	0.017	0.892

Table 11 Difference-in-differences results

Dependent variable:		With matching method	
		PSM	CEM
Deposit rate	Without matching method	Set 3	Set 3
<i>Panel A. With common trends</i>			
Treated $\square$ $\times$ Post	−0.0106**	−0.0104**	−0.0029
Controls	Y	Y	Y
Fixed effects	Y	Y	Y
Cluster SE	Bank	Bank	Bank
Number of observations	85,948	72,494	10,891
Number of banks	6,471	6,466	1,148
R-squared	0.923	0.921	0.953
<i>Panel B. Without a common trend</i>			
Treated $\square$ $\times$ Post	−0.0338***	−0.0209***	−0.0118
Controls	Y	Y	Y
Fixed effects	Y	Y	Y
Cluster SE	Bank	Bank	Bank
Number of observations	35,766	30,511	3,132
Number of banks	3,501	3,487	457
R-squared	0.916	0.914	0.954

Note: * denotes significance at 10%, ** at 5%, and *** at 1%

a common trend, while Panel B shows the results without a common trend. As can be seen in Panel A, branches in counties struck by natural disasters experience a significant negative impact on deposit rates as compared to branches in unaffected adjacent counties with and without matching (Set 3) based on PSM. However, when matching with CEM using the same variables (Set 3), the results are quite different. In this case, there is no significant effect of natural disasters on deposit rates. In other words, the matching method matters when the common trends assumption is satisfied.

One concern is that CEM has a smaller sample of matched observations. However, this is not a critical problem because a smaller sample size might indicate that (1) more observations are dropped since well-matched units rarely exist, and (2) heterogeneity between the treatment and control groups is considerably reduced by matching and eliminating observations. This leads to smaller variances in causal estimates and, in turn, a low imbalance measure (Iacus et al. 2011, 2012). Indeed, in observational data, MIB methods may even produce no matched observations (Iacus et al. 2011).

Next, the same type of analysis is performed without a common trend assumption satisfied. The results, in this case, are reported in Panel B and are essentially the same as those in Panel A. Branches in counties struck by natural disasters experience a significant negative impact on deposit rates as compared to branches in unaffected adjacent counties with and without matching (Set 3) based on PSM. However, when matching with CEM using the same variables (Set 3), natural disasters have no significant effect on deposit rates. Thus, once again, the matching method matters even when a common trend assumption is not satisfied. Moreover, the failure of this assumption to be satisfied does not affect the empirical results.

Table 12 summarizes our findings for the two different matching methods, PSM and CEM, using the same set of variables (Set 3) with and without a pretreatment common trend. Importantly, the results are quite different for PSM and CEM. In the former case, deposit rates are significantly affected by natural disasters, whereas in the latter case the rates are not affected. However, the lack of a common trend in the treatment and control groups does not affect the empirical results. In short, matching and the matching method employed matter, but a violation of a common trend assumption does not matter for our sample of branches and study period.

Table 12 Summary of results

<i>HIGH generalizability</i>				
Δ Imbalance measure		Matching methods		
		No matching	PSM	CEM
			Set 3	Set 3
			Low	High
Common trends	Not satisfied	Significant	Significant	Not significant
	Satisfied	Significant	Significant	Not significant

77.7 Conclusions

The contribution of our entry has been twofold. First, we analyzed the effect of natural disasters on the deposit rates of banks. This was done using the DID identification strategy. The basic (i.e., most reliable) finding is that there is no statistically significant response of deposit rates by branches of banks located in counties struck by natural disasters as compared to deposit rates of branches of banks in adjacent counties not so affected. Of course, our focus was only on the rate for 12-month certificates of deposit (CDs) requiring a minimum balance of \$10,000. This finding is consistent with that of Dlugosz et al. (2020), who also use this rate in their analysis and find that there is no statistically significant effect of natural disasters on the rate for branches in counties experiencing them as compared to branches in adjacent counties not experiencing them.¹² However, in their triple DID analysis they find a statistically significant effect for those branches that set rates within the county. More generally, our results are not directly comparable to their results because, as Angrist and Pischke (2010, p. 23) state, “Empirical evidence on any given causal effect is always local, derived from a particular time, place, and research design.”

Second, as noted earlier, there are several steps in conducting our analysis where discretion in how to proceed could be exercised. This involved decisions about whether to match or not and what method to use if matching. Also, a decision had to be made as to whether there was a common trend in the outcome variable before the treatment for both the treatment group and the control group. Our findings indicate that the choices made do indeed affect the empirical results. If we do not match, the results indicate that natural disasters have a statistically negative effect on deposit rates without matching and with a low degree of matching. However, when we use a medium or a high degree of matching there is no statistically significant effect. Moreover, when we use two different matching methods, the results differ. In the case of PSM, we find a statistically significant effect, but no effect in the case of CEM. Lastly, the results do not change whether or not there is a common trend for the treatment group and control group.

Overall, the findings are important not only in terms of assessing the effect of natural disasters on deposit rates of bank branches but in terms of showing that the discretion a researcher possesses can lead to quite different results, whether it be significant or insignificant results, and thereby possibly a publication or not. More importantly, studies with policy implications can be misleading depending on the choices a researcher makes.

¹²Dlugosz et al. (2020) actually use the CD rate following the disaster in the event window minus the rate in the period prior to the disaster in the event window.

Appendices

Appendix 1

The multivariate imbalance measure is calculated as below (Blackwell et al. 2009; Iacus et al. 2011, 2012):

$$\mathcal{L}_1(f, g) = \frac{1}{2} \sum_{\ell_1, \dots, \ell_k \in H(X)} |f_{\ell_1, \dots, \ell_k} - g_{\ell_1, \dots, \ell_k}|$$

Iacus et al. (2012) let $f_{\ell_1, \dots, \ell_k}$ and $g_{\ell_1, \dots, \ell_k}$ denote the k -dimensional relative frequencies of the treatment and the control units, respectively, affiliated with coordinates $\ell_1, \dots, \ell_k$. $H(X)$ denotes the values generated by the multidimensional bins (H) chosen for variable X . f^m and g^m denote the frequencies of matched treatment and control units, respectively, corresponding with the frequencies of nonmatched f and g . Then, when the matching is conducted, a good matching method produces a reduction in the imbalance measure; $\mathcal{L}_1(f^m, g^m) \leq \mathcal{L}_1(f, g)$ (Blackwell et al. 2009; Iacus et al. 2011, 2012).

Appendix 2

Average treatment effect on the treated (ATT), α_{ATT} , of the DID approach is:

$$\alpha_{ATT} = E(Y_t^1 | D = 1, X) - E(Y_t^0 | D = 1, X)$$

where D , Y_t^1 , and Y_t^0 denote the treatment indicator (0 or 1), the potential outcome *with* treatment, and the potential outcome *without* treatment, respectively. $t = 0$ indicates pretreatment period and 1 indicates posttreatment period. The common trend assumption is expressed as follows (Lechner 2011):

$$E(Y_1^0 | D = 1, X) - E(Y_0^0 | D = 1, X) = E(Y_1^0 | D = 0, X) - E(Y_0^0 | D = 0, X)$$

This equation indicates that if there were no treatment, the treated and untreated groups ($D = 1$ and $D = 0$) would show the same time trend given observed covariates X , which are assumed not to vary over time (Lechner 2011). Thus, the difference in the trend between the observed outcomes with treatment and those without treatment would be attributed to the treatment effect, not to the differences in any covariate between the treatment and control groups (Lechner 2011). This is the main idea of the DID approach to attain identification by calculating the difference between the average outcome of treatment and control groups after the treatment and subtracting the difference of the outcomes that existed before the treatment (Lechner 2011).

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

- . 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives* 24 (2): 3–30.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. How much should we trust differences-in-differences estimates?. *The Quarterly journal of economics* 119 (1): 249–275.
- Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro. 2009. cem: Coarsened exact matching in Stata. *The Stata Journal* 9 (4): 524–546.
- Brodeur, Abel, Nikolai Cook, and Anthony Heyes. 2020. Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review* 110 (11): 3634–3660.
- Cochran, William G., and Donald B. Rubin. 1973. Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A* 35 (4): 417–446.
- Diamond, Alexis, and Jasjeet S. Sekhon. 2013. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics* 95 (3): 932–945.
- Đlugosz, Jennifer, Yong Kyu Gam, Radhakrishnan Gopalan, and Janis Skrastins. 2020. Decision-making delegation in banks. Available at SSRN 3155683.
- Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. Characterizing selection bias using experimental data. *Econometrica* 66 (5): 1017–1098.
- Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15 (3): 199–236.
- Iacus, Stefano M., Gary King, and Giuseppe Porro. 2011. Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association* 106 (493): 345–361.
- . 2012. Causal inference without balance checking: Coarsened exact matching. *Political Analysis* 20 (1): 1–24.
- Imai, Kosuke, and Marc Ratkovic. 2014. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B: Statistical Methodology* 76: 243–263.
- Imai, Kosuke, Gary King, and Elizabeth A. Stuart. 2008. Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171 (2): 481–502.
- King, Gary, and Richard Alexander Nielsen. 2019. Why propensity scores should not be used for matching. *Political Analysis* 27: 4.
- Lechner, Michael. 2011. *The estimation of causal effects by difference-in-difference methods*. Now.
- Li, Mingxiang. 2013. Using the propensity score method to estimate causal effects: A review and practical guide. *Organizational Research Methods* 16 (2): 188–226.
- Mora, Ricardo, and Iliana Reggio. 2015. didq: A command for treatment-effect estimation under alternative assumptions. *The Stata Journal* 15 (3): 796–808.
- . 2019. Alternative diff-in-diffs estimators with several pretreatment periods. *Econometric Reviews* 38 (5): 465–486.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1): 41–55.
- Rubin, Donald B. 1973. The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*: 185–203.
- Rubin, Donald B., and Elizabeth A. Stuart. 2006. Affinely invariant matching methods with discriminant mixtures of proportional ellipsoidally symmetric distributions. *The Annals of Statistics* 34 (4): 1814–1826.
- Rubin, Donald B., and Neal Thomas. 1992. Affinely invariant matching methods with ellipsoidal distributions. *The Annals of Statistics*: 1079–1093.
- . 2000. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association* 95 (450): 573–585.
- Stuart, Elizabeth A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 25 (1): 1.

Financial Panel Data Models, Strict Versus Contemporaneous Exogeneity, and Durbin-Wu-Hausman Specification Tests

78

Robert H. Patrick

Abstract

Panel data, pooling cross-sectional and time series data, is increasingly used in estimating financial models. This chapter presents estimators for pooled, fixed effects (FE), and random effects (RE) linear panel data models, assumptions on which they are based, particularly contemporaneous and strict exogeneity assumptions on the explanatory variables, and model specification testing (DWH and Wooldridge tests). Implications of estimator and specification choice on parameter consistency and standard error efficiency are developed. The relationship between the approaches, choice of approach, and their advantages and disadvantages are discussed. Examples illustrate application of the estimators and tests, applying the concepts and testing specifications to illustrate their use and interpretation.

Keywords

Panel data · Fixed effects · Random effects · Endogeneity, Strict exogeneity · Specification tests

JEL Classification

C58 · C23 · C51 · C52

78.1 Introduction

Omitted variables are a common cause of endogeneity in estimation of empirical finance models (e.g., Giglio and Xiu 2021). Panel data, combining time series on each cross-sectional entity in the sample, presents an opportunity to model unobserved

R. H. Patrick (✉)

Professor Emeritus, Finance and Economics, Rutgers University, Newark, NJ, USA

e-mail: rpatrick@business.rutgers.edu

heterogeneity that may exist across cross section and time, potentially providing approaches to obtain unbiased and/or consistent parameter estimates. Consider CEO compensation for example. An observed variable such as compensation may be correlated with an unobserved (to the econometrician) variable such as effort or ability, which implicitly enters the model through an individual specific effect. Panel data approaches allow us to model such unobserved heterogeneity and reduce or eliminate omitted variables problems. Chang and Lee (1977) provide an early example.

Using panel data, rather than only cross-sectional or time series data, provides a number of additional potential benefits, including a larger sample size (which is also an approach to ameliorating multicollinearity problems and to increase the precision of our estimates). Panel data also allow consideration of two dimensions, time series and cross section, providing greater flexibility in identifying the data generating process (DGP). An example is to distinguish technological change, a time series notion, from economies of scale, a cross-sectional notion. A cross section might inaccurately estimate the effect of increasing capital, while panel time series components can be used to incorporate the impact of technical change. Finally, we can also relax constant parameter assumptions using panel data estimation techniques (as simply as by interacting effects with explanatory variables).

Often maintained, but rarely tested in financial panel data models are the contemporaneous and strict exogeneity assumptions.¹ Violations of either or both of these assumptions imply parameter estimates will be inconsistent, standard errors will be inefficient, and the usual hypothesis tests will no longer be valid. The CEO compensation example above is a contemporaneous issue. In addition, the strict exogeneity concern in financial models is primarily about potential correlation between a current dependent variable and future values of explanatory variables. According to Grieser and Hadlock (2019), the majority of published research applying panel modeling techniques in the top three finance journals use the fixed effects model and only one article (out of 216) provided any test of strict exogeneity.² Although violation of contemporaneous and strict exogeneity may be prevalent in financial panel data estimation, these issues with consistent parameter estimation have largely been implicitly assumed away. This is not to imply that these papers have presented inconsistent estimates, but these assumptions can be tested rather than maintained when there is any question regarding the endogeneity of any of the explanatory variables.

In this chapter, the basic approaches to linear (in parameters) panel data model specification, estimation, and testing are introduced, and examples of their application are provided to show how results are interpreted.³ First, we define exogeneity, strict

¹Any panel data model with a lagged dependent variable as an explanatory variable and cross-sectional unobserved effects will by specification violate strict exogeneity.

²The three journals Grieser and Hadlock (2019) searched were the *Journal of Finance*, *Journal of Financial Economics*, and the *Review of Financial Studies* from 2006 to 2013, with an updated search of the *Journal of Finance* for 2017.

³Linear is used here to imply that parameters to be estimated enter the model linearly, rather than in a nonlinear fashion. The techniques covered here are applicable to nonlinear in variables models, but not nonlinear in parameters models.

exogeneity, and endogeneity, and discuss their implications for unbiased and consistent parameter estimation in the next section. Sections that follow then provide the fundamental assumptions and considerations in estimating and specification testing panel data models. Pooled least squares (LS), fixed effects (FE), and random effects (RE) are considered. Sections following the specification and estimation of panel models provide tests for contemporaneous and strict exogeneity. Durbin-Wu-Hausman (DWH) type tests can accomplish both, but Wooldridge's (2010) strict exogeneity tests are also discussed and applied. Empirical applications are then provided illustrating the modeling issues, valid application, and interpretation of DWH tests, and estimation approaches to address contemporaneous and strict endogeneity issues.

78.2 Exogeneity, Strict Exogeneity, and Endogeneity

Consider a linear (in parameters) panel model

$$y_{it} = X_{it}\beta + \varepsilon_{it}, i = 1, \dots, N, t = 1, \dots, T \quad (1)$$

and

$$\varepsilon_{it} = u_t + v_i + e_{it} \quad (2)$$

where X_{it} is a $1 \times k$ vector of explanatory variables, β is a $k \times 1$ vector of parameters, ε_{it} is the composite error potentially comprised of components u_t for the time effect t , v_i for the cross-sectional effect i , and the common error, e_{it} , that affect cross-section i at time t . It is generally assumed that the u_t , v_i , and e_{it} are independent across all t , i , and it , respectively.

There are a number of potential violations with the specified model. Constant parameters may be overly restrictive, perhaps they should be specified as random variables, or a nonlinear in parameters model may better explain the data generating process. Unless we are working with experimental data (as in Example 1 below), we cannot generally expect all the explanatory variables to be deterministic. In fact we should not expect this with the usual type of data used in most empirical finance and economics models. The issue is that stochastic explanatory variables may be related to the errors, leading to parameter bias and inconsistency.

Any explanatory variable in the model that is contemporaneously correlated with the error is termed endogenous, regardless of the source of the correlation. Sources of endogeneity in econometrics are typically omitted variables, measurement error, simultaneity, and of selection. If there is no correlation between the explanatory variable and the error at any time, then the explanatory variable is (contemporaneously) exogenous. If there is no correlation between the explanatory variable and the error in any other period, $t = 1, \dots, T$, then the variable is strictly exogenous.⁴

⁴See Engle, Hendry, and Richard (1983) for additional detail. Often exogenous is used to imply strict exogeneity (e.g., Davidson and MacKinnon 2004).

More formally, contemporaneous exogeneity implies

$$E(e_{it}|X_{it}, u_t, v_i) = 0 \quad \forall i = 1, \dots, N; t = 1, \dots, T, \quad (3)$$

i.e., the error for cross-section i at time t is not correlated with any contemporaneous explanatory variable or effect. If only the parameters associated with the X explanatory variables are of interest with can drop the cross-sectional and time effects from the conditional. A stronger assumption than (3) is strict exogeneity, which implies

$$E(e_{it}|X_{is}, u_t, v_i) = 0 \quad \forall i = 1, \dots, N, s, t = 1, \dots, T, \quad (4)$$

or, more succinctly in vector/matrix form,

$$E(e|X, u, v) = 0. \quad (5)$$

That is, the error at each point in time is not correlated with any contemporaneous, lead, or lag of any of the explanatory variables, for all observations. This is usually implicitly maintained. This latter assumption, (4), may be violated when, e.g., high (or low) realizations of the dependent variable (e.g., firm performance) at time t have a positive or negative impact on future values of an explanatory variable (e.g., managerial ownership). Correlation between the current y_{it} and future values of explanatory variables is a potentially prevalent strict exogeneity issue in financial panel data modeling that is critical to consistent parameter estimation in these models.⁵

78.3 Linear Panel Data Models and Pooled LS Estimation

Consider the linear (in parameters) population model

$$y_{it} = \beta_0 + X_{it}\beta + e_{it}, i = 1, \dots, N, t = 1, \dots, T \quad (6)$$

where β_0 is the common constant term, X_{it} is a $1 \times k$ vector of explanatory variables, β is a $k \times 1$ vector of parameters associated with the explanatory variables, and e_{it} is the error. Stacking each cross section of time series data and including a common constant term leads to

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{NT} \end{pmatrix} = \begin{pmatrix} 1 & x_{111} & \cdots & x_{k11} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1NT} & \cdots & x_{kNT} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} e_{11} \\ \vdots \\ e_{NT} \end{pmatrix}. \quad (7)$$

⁵Strict exogeneity may also be an issue in time series models.

Given that the data is ordered so that all observations of the first cross section appear first, etc. for $i = 1, \dots, N$, then the row corresponding to observation it is row $T(i - 1) + t$. Equation (7) implies that there is a common constant term across all cross sections through time, a critical difference in pooled LS compared to FE and RE models. In matrix notation, keeping the constant and slope parameters separate, (7) can be expressed more concisely in vector/matrix notation as

$$y = \beta_0 + X\beta + e, \quad (8)$$

where y and e are NT vectors, X is a $NT \times k$ matrix, β_0 is a constant term common to the entire sample, and β is a $k \times 1$ vector of slope parameters.

The analysis presented here is for a random sample of cross sections of data (e.g., firms, portfolios, individuals, etc.), $i = 1, \dots, N$, over time, $t = 1, \dots, T$, providing NT total observations.⁶ It is assumed for now that the panel is balanced, i.e., there are the same number of time series observations for all cross sections in the sample. Unbalanced panel data is addressed in Sect. 78.7. The econometric methods presented here, unless otherwise stated, are for panel data with a large number of cross sections relative to the number of time series per cross section, i.e., $N > T$. For panel data models, the cross section is assumed to be a random sample, as with cross-sectional data, but, since we should generally expect variables to be correlated over time, correlation across time is relatively unrestricted. Random implies we can collect a sample of independent and identically distributed cross sections of the underlying population of interest.

In addition to the population model and random sampling assumptions above, we need to estimate the parameters and their standard errors. There are four primary pooled LS assumptions that are key to consistent estimation of parameters and efficient estimation of standard errors. These assumptions are also stated in terms of the population model. Some of these pooled LS assumptions apply and some need to be modified for the FE and RE models later in this chapter.

Assumption P1: Strict exogeneity, $E(e_{it}|X_{is}) = 0$ for all i, s , and t .

Another way to state this assumption that may be more informative in the context of financial panel data modeling is

$$E(y_{it}|x_{i1}, \dots, x_{iT}) = E(y_{it}|X_{it}) = \beta_0 + X_{it}\beta, \text{ for } t = 1, \dots, T. \quad (9)$$

because it is easy to see that strict exogeneity implies that, controlling for X_{it} , X_{is} has no effect on y_{it} for all $s \neq t$, $s, t = 1, \dots, T$.

The key assumption for consistent pooled LS parameter estimation is the population moment (or orthogonality) condition

⁶If the error term for cross-sect. i , conditional on the explanatory variables, is independent of being selected into the sample, random can be relaxed (random is actually a stronger assumption than is needed).

$$E(X'e) = 0, \quad (10)$$

i.e., the error has zero mean and is uncorrelated with each of the explanatory variables in the population. If X includes a constant, let $X_1 = 1$, then (10) implies

$$E(X_1e) = E(e) = 0 \text{ and } \text{cov}(X_j, e) = 0 \text{ for all } j = 2, \dots, K. \quad (11)$$

Sufficient but not necessary for (10) and (11) is the zero conditional mean assumption

$$E(e|X) = 0, \quad (12)$$

and the expected error term has a conditional mean of zero. This implies that the unconditional mean of the error term is zero, $E(e) = 0$ (a special case of the Law of Iterated Expectations), and that e is not correlated with any X or function of any X at any time, $t = 1, \dots, T$, the function of being the much stronger of the two.

Example: Suppose our population model is

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + e, \quad E(e|x_2, x_3) = 0, \quad (13)$$

then

$$E(e) = 0, \text{cov}(X_2, e) = 0, \text{cov}(X_3, e) = 0. \quad (14)$$

Equation (13) includes the strict exogeneity assumption, which implies that e is uncorrelated with any other function of the X variables, e.g., X_2X_3 , X_2^2 , X_3^2 , etc., so that all effects of X_2 and X_3 on the expected value of y are included in estimating the parameters. If (13) holds, then we have consistently measured the partial effects of a variable when we differentiate y with respect to an X variable. For example, if strict exogeneity holds, $E(e|X) = 0$, with the linear in parameters model then, e.g., $\partial E(y|X)/\partial X_2 = \beta_2$. That is, we have an appropriate functional form specification conditional on the X variables included in the specified model. Equation (13) is a stronger assumption than the contemporaneous correlation represented in (14), the latter leaves open the possibility that e is correlated with nonlinear functions of the variables.

Assumption P1 and the additional pooled LS assumptions below assure that we can identify and consistently estimate the pooled model parameters, with efficient standard errors.

Assumption P2: Explanatory variables are nonstochastic or bounded, i.e., the trace norm of X_{it} , $\|X_{it}\| < K < \infty$ or satisfy moment conditions $E\left\|(x_{it} - \bar{x})(x_{js} - \bar{x})'\right\| < K < \infty \forall i, j, s, t$, where $\bar{x} = \sum_{t=1}^T \sum_{N=1}^N x_{it}/NT$. This assumption can be relaxed to include trend or unit root processes. Explanatory variables that are nonstochastic or deterministic will naturally be strictly exogenous.

Assumption P3: The $k \times k$ matrix

$$Q_{P,NT} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x})(x_{it} - \bar{x})' \quad (15)$$

is positive definite for all N and T as $N \rightarrow \infty$.⁷ Since $E(X'X)$ is a $k \times k$ symmetric matrix, this is the same as assuming that the rank $[E(X'X)] = k$. Perfect multicollinearity violates this condition. To avoid a violation of this assumption, eliminate any linear combinations of data in X .

Assumption P4: The errors e_{it} are independent cross-sectionally and may be heteroskedastic cross-sectionally and correlated across time,

$$E(e_{it}^2|X) = \sigma_i^2 < K < \infty, \quad (16)$$

$$E(e_{it}e_{js}|X) = 0 \text{ if } i \neq j \forall s, t = 1, \dots, T, \quad (17)$$

and

$$E(e_{it}e_{is}|X) = \sigma_i(s, t), \text{ for } s \neq t, \quad (18)$$

where the bounded error process e_{it} has autocovariances $\sigma_i(s, t)$. If T is fixed and $N > T$, then restrictions on the autocovariances are not necessary.

If $E(e_{it}|X_{it}) = 0$ for all $i = 1, \dots, N$ and $t = 1, \dots, T$ (i.e., assumption P1 holds), then (6) can be estimated by LS, which will be efficient if the e_{it} are independently and identically distributed. The pooled LS parameter estimates are obtained by minimizing the sum of squared errors, $e'e$, to find

$$\hat{\beta}_{LS} = \left[\sum_{t=1}^T \sum_{i=1}^N (X_{it} - \bar{x})(X_{it} - \bar{x})' \right]^{-1} \left[\sum_{t=1}^T \sum_{i=1}^N (X_{it} - \bar{x})(y_{it} - \bar{y})' \right], \quad (19)$$

where the sample means are

$$\bar{y} = \sum_{t=1}^T \sum_{i=1}^N y_{it}/NT \text{ and } \bar{x} = \sum_{t=1}^T \sum_{i=1}^N x_{it}/NT.$$

Using (15), the variance of $\hat{\beta}_{LS}$ is given by

$$\text{var}(\hat{\beta}_{LS}|X) = Q_{P,NT}^{-1} \left[\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T E(e_{it}e_{js}|X)(x_{it} - \bar{x})(x_{jt} - \bar{x})' \right] Q_{P,NT}^{-1} \quad (20)$$

⁷More precisely, as N and/or $T \rightarrow \infty$.

Applying Assumption P4 to substitute for $E(e_{it}e_{js}|X)$ in (20), the pooled LS variance can be expressed as

$$\begin{aligned} \text{var}(\hat{\beta}_{LS}|X) &= Q_{P,NT}^{-1} \left[\sum_{i=1}^N \sum_{t=1}^T \sigma_i^2 (x_{it} - \bar{x})(x_{it} - \bar{x})' \right] Q_{P,NT}^{-1} \\ &\quad + Q_{P,NT}^{-1} \left[\sum_{i=1}^N \sum_{t \neq s}^T \sigma_i(s, t) (x_{it} - \bar{x})(x_{it} - \bar{x})' \right] Q_{P,NT}^{-1}, \end{aligned} \quad (21)$$

where $\sigma_i(s, t)$ are the autocovariances for the pooled LS estimator. Although the pooled LS parameter estimates may be consistent, the standard errors will be inefficient and likely underestimated. See Pesaran (2015) for additional detail.

β is identified with Assumptions P1 and P3 in the population model (6).

Consistency implies that, if our sample is large enough, the estimator produces an estimate that is close to the population parameter with high probability. Precisely, with a sample of size N , let $\hat{\beta}_N$ be an estimator of the population β . Then $\hat{\beta}$ is a consistent estimator of β if

$$\lim_{N \rightarrow \infty} P(|\hat{\beta}_N - \beta| < \varepsilon) = 1, \quad (22)$$

where ε is positive and arbitrarily small. So the probability that $\hat{\beta}_N$ falls in the interval $[\beta - \varepsilon, \beta + \varepsilon]$ can be made arbitrarily close to 1 with a sufficiently large sample size, regardless of the smallness of ε . If (22) is true, then the sequence of random variables $\hat{\beta}_N$ converges in probability to β and the probability limit of $\hat{\beta}_N$, abbreviated as

$$\text{plim}(\hat{\beta}) = \beta.$$

To summarize consistency of pooled LS: given a random sample of the population (6), Assumptions P1 and P3 imply that $\text{plim}(\hat{\beta}) = \beta$, i.e., $\hat{\beta}$ is a consistent estimator of β . See Greene (2018) and Pesaran (2015) for additional detail. Under assumptions P1 and P3, pooled LS parameters are identified and parameter estimates are unbiased and consistent if $E(e|X) = 0$, that is, if the X are strictly exogenous. If $E(e|X) \neq 0$, then pooled LS is biased and inconsistent. This will be the case if, e.g., X_{it} contains lagged dependent variables and the e_{it} are autocorrelated, or if y_{it} is correlated with future values of any of the explanatory variables.

Concerning efficiency of the estimator, it is unlikely that the errors are independently and identically distributed given panel data is being analyzed because shocks affecting cross-section i for all $t = 1, \dots, T$ imply correlation between e_{it} and e_{is} for all $t \neq s$, and shocks affecting all cross-sections $i = 1, \dots, N$ at a particular t imply correlation between e_{it} and e_{jt} for all $i \neq j$. So even if strict exogeneity holds and we have consistent parameter estimates, LS applied to panel data (the pooled LS estimator) is likely to produce inefficient standard errors (generally too small) due

to an inconsistent estimate of the covariance matrix. See Example 1 below for an illustration and discussion of these points.

78.4 Unobservable Heterogeneity

To estimate the panel data model, the unobserved heterogeneity, the u_t and v_i , are specified as FE, RE, or mixed. So the effects are captured as explanatory variables or error components. FE parameters on the X variables can be estimated using group dummy variables or by group mean differencing the data and applying least squares (LS). Including the group dummy variables as X variables is required to estimate parameters on the effects. If effects are specified as random, the random effects covariance matrix is specified as functions of the variances of the u_t and/or v_i , and e_{it} . Feasible generalized least squares (FGLS), maximum likelihood (ML), or generalized method of moments (GMM) could be used for estimation.

Joint hypothesis tests of the fixed effects can be carried out with an F-test. For random effects, a large sample test is required so a Wald, Likelihood Ratio (LR), or Lagrange multiplier (LM) test could be applied.

A Durbin, Wu, Hausman (DWH) test can be used to choose between the FE or RE specification of the panel model. If the effects are correlated with the explanatory variables, then RE will be inconsistent. Another form of DWH test can be used to test strict exogeneity, as can some other tests developed in Wooldridge (2010).

78.5 Fixed Effects (FE) Models

While empirical examples below will contain both cross section and time effects, we will mostly abstract from the time effects in the theoretical development. So to simplify the exposition, let $u_t = 0$ in the error components specification (2). This implies that we are considering shocks specific to the cross section but no time-specific shocks for now. The model is then

$$y = X\beta + D\gamma + e = X\beta + \sum_{i=1}^N \gamma_i D_i + e, E(ee') = \sigma_e^2 I_{NT}, \quad (23)$$

where y and e are NT vectors and D is a $NT \times N$ matrix of dummy variables such that the element in the row corresponding to observation it , for $i = 1, \dots, N$ and $t = 1, \dots, T$, and column j , for $j = 1, \dots, N$, is equal to 1 if $i = j$ and 0 otherwise. The parameter vector γ measures the affects of the *fixed effects*, v_i . The N vector γ has typical elements v_i since the NT vector $D\gamma$ has v_i in the row corresponding to observation it . Therefore there is exactly one element of D equal to 1 in each row, implying that the NT vector i with each element equal to 1 is a linear combination of the columns of D . They may be correlated with the X variables but have to be independent of the errors e_{it} .

The FE assumptions for consistent β parameter estimation with efficient standard errors are the same as the pooled LS assumptions P1 and P4, but assumptions P2 and P3 must be revised to account for the panel nature of the data as follows.

Assumption FE2 Explanatory variables are nonstochastic or bounded, i.e., the trace norm of X_{it} , $\|X_{it}\| < K < \infty$ or satisfy moment conditions $E\left\|(x_{it} - \bar{x}_i)(x_{js} - \bar{x}_j)'\right\| < K < \infty \forall i, j, s, t$, where $\bar{x}_i = \sum_{t=1}^T x_{it}/T$.

Assumption FE3

$$Q_{FE,NT} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \quad (24)$$

is a $k \times k$ positive definite matrix for all N and T as $N \rightarrow \infty$.

If the parameters associated with the fixed effects are of interest, consistent estimation of the fixed effects parameters associated with v_i in the FE model, pooled LS assumption P1 would be revised as follows.

Assumption FE1 Strict exogeneity, $E(e_{it}|X_i, v_i) = 0$, for all i, s , and t .

The LS estimator for the FE model is unbiased if X and D are strictly exogenous (if there is no contemporaneous correlation between explanatory variables and the error term). If these assumptions hold, the FE estimator is BLUE (i.e., the Gauss-Markov Theorem holds). Usually these parameters are treated as nuisance parameters. Consistent estimation of the parameters associated with the explanatory variables X is usually the research interest.

The fixed effects model (23), conditionally on γ , satisfy these moment conditions:

$$E(X'_{it}(y_{it} - X_{it}\beta - v_i)) = 0 \text{ and } E(y_{it} - X_{it}\beta - v_i) = 0. \quad (25)$$

From Frisch and Waugh (1933) and Lovell (1963), we have what is known as the Frisch-Waugh-Lovell Theorem, which demonstrate that the FE estimates of β in (23) can be obtained using LS to regress $M_D y$, the errors from a regression of y on D , on $M_D X$, the matrix of errors from regressing each column of X on D . For any NT vector x , the cross-sectional or group mean is $\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}$. A straightforward check of element it of the matrix $M_D X$ reveals that it is equal to the deviation from the group mean, $x_{it} - \bar{x}_i$. That is, the fixed effects estimates of parameters, β , associated with the explanatory variables X , can be obtained by transforming the data by group-mean differencing, and using the pooled LS estimator in the following regression:

$$\left(y_{it} - \sum_{t=1}^T \frac{y_{it}}{T}\right) = \sum_{h=1}^K \beta_j \left(x_{hit} - \sum_{t=1}^T \frac{x_{hit}}{T}\right) + \left(e_{it} - \sum_{t=1}^T \frac{e_{it}}{T}\right). \quad (26)$$

In the linear (in parameters) FE model, the sufficient statistic is the group mean (a statistic is called sufficient for a parameter if the distribution of the sample given i does not depend on the parameter). This is referred to as the within-groups estimator because it uses only information on the variation around the mean of each of the N groups. The LS FE parameter estimates are then given by

$$\begin{aligned}\hat{\beta}_{FE} &= \left[\sum_{t=1}^T \sum_{i=1}^N (X_{it} - \bar{x}_i)(X_{it} - \bar{x}_i)' \right]^{-1} \left[\sum_{t=1}^T \sum_{i=1}^N (X_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)' \right] \\ &= \left(\sum_{i=1}^N X_i' M_T X_i \right)^{-1} \left(\sum_{i=1}^N X_i' M_T y_i \right)\end{aligned}\quad (27)$$

where the group (cross section) means are $\bar{y}_i = \sum_{t=1}^T y_{it}/T$ and $\bar{x}_i = \sum_{t=1}^T x_{it}/T$, and M_T is a projection matrix defined analogously to M_D above. The fixed effect parameters γ_i can be estimated using the specified model with (27) and the FE assumptions,

$$\hat{\gamma}_i = \bar{y}_i - \bar{\beta}_{FE}' \bar{x}_i. \quad (28)$$

Using assumptions P1 and FE3, it is straightforward to show that the FE LS estimator of $\hat{\beta}$ is unbiased, i.e., $E(\hat{\beta}_{FE}) = \beta$.

The variance of $\hat{\beta}_{FE}$ is given by

$$\text{var}(\hat{\beta}_{FE}|X) = (1/NT) Q_{FE,NT}^{-1} \left[\frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N X_i' M_T E(e_i e_j' | X) M_T X_j \right] Q_{FE,NT}^{-1}. \quad (29)$$

Equation (29) can be simplified further using assumption P4 to substitute for $E(e_i e_j' | X)$. See Pesaran (2015) for details and the derivation of a number of other results in the literature that can be derived from here.⁸

An advantage of the FE estimator that has made it so popular is that it is relatively easy to compute. If parameters associated with the dummy variables D are not of interest, then there is no need to estimate the fixed effects parameters, all that is needed is to compute the N group means for all variables and transform the data by subtracting the appropriate mean from the observation. That is, specification of the dummy variables for the fixed effects is required only if estimates of $\hat{\gamma}$ are of interest. Multicollinearity may be an issue in precisely separating the individual effects. If the X variables are well explained by the dummy variables in D , then β will not be

⁸These other results in the literature are beyond the scope of this chapter so are not covered further here. Pesaran (2015), Chamberlain (1984), Cameron and Trivedi (2005), Wooldridge (2010), Hsiao (2014), Greene (2018), and Baltagi (2021) provide detail.

precisely estimated with FE, due to the correlation between X and D . Exact multicollinearity prohibits some specifications such as those with explanatory variables that are constant for each cross section since such a variable would be perfectly collinear with the columns of dummy variables, D (see Application 1 below for an example).

78.6 Random Effects (RE) Models

The efficiency of the FE estimator can be improved by applying the RE estimator, but only if the X and v_i are independent. As discussed above, what is required with the FE model is that the e_{it} are independent of X and v_i (and u_t if time effects are also specified in the model). Contrary to the FE model, RE allows estimation of parameters associated with variables that are fixed across the cross section, as discussed in the previous section of this chapter.

The RE assumptions for consistent parameter estimation with efficient standard errors are:

Assumption RE.1 Strict exogeneity implies the error is conditional on the explanatory variables X and the effects, $E(e_{it}|X_i, v_i) = 0$, and the random effects are constant through time and independent of the explanatory variables X_i , $E(v_i|X_i) = 0$, for all i and t .

Assumption RE.2 $E(e_i e_i' | X_i, v_i) = \sigma_e^2 \mathbf{I}_T$ and $E(v_i^2 | X_i) = \sigma_v^2$, for all $i = 1, \dots, N$.

Assumption RE.3 $\sum_{i=1}^N X_i' \Sigma^{-1} X_i$ is a nonsingular matrix for all N and T as $N \rightarrow \infty$.

RE is not consistent if the random effects are correlated with the X variables. For example, consider a panel of observations on CEO pay, an observed variable like compensation may be correlated with an unobserved (to the econometrician) variable such as effort or ability, which implicitly enters the individual specific error v_i . If this individual effect, an error component with RE, is correlated with variables explaining CEO compensation, then their associated parameter estimates will be inconsistent.

If X and v_i are independent then

$$E(e_{it}|X_{it}) = E(v_i + e_{it}|X_{it}) = 0 \quad (30)$$

since e_{it} is also independent of X . Condition (30) implies that the pooled LS model, (6), provides unbiased parameter estimates (see Example 1 below). However, LS estimation of the pooled panel model is unlikely to be efficient because the e_{it} are not independently and identically distributed. The covariance matrix of the e_{it} can be computed if the v_i are $iid(0, \sigma_v^2)$, so the name “random effects.” From the error

components, $\varepsilon_{it} = u_t + v_i + e_{it}$, assuming shocks are independent for simplicity and $u_t = 0$, it follows that

$$\begin{aligned}\text{Var}(\varepsilon_{it}) &= \sigma_v^2 + \sigma_e^2 \\ \text{Cov}(\varepsilon_{it}\varepsilon_{jt}) &= \sigma_v^2 \text{ for all } t \text{ and } i \neq j, \\ \text{Cov}(\varepsilon_{it}\varepsilon_{js}) &= 0 \text{ for all } i, j \text{ and } t \neq s,\end{aligned}\quad (31)$$

define the elements of the $NT \times NT$ covariance matrix, Ω , needed for GLS estimation. Assumptions in (31) allow heteroskedasticity and contemporaneous correlation between the errors across the cross sections, but rule out autocorrelation, i.e., there are no lagged correlations between the errors. If the data are ordered by the N cross-sectional units in groups of T observations each (assuming a balanced panel), the matrix has the form

$$\Omega = E(ee') = I_N \otimes \Sigma = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma_N \end{bmatrix} \quad (32)$$

where

$$\Sigma_i \equiv \begin{bmatrix} \sigma_e^2 + \sigma_v^2 & \sigma_v^2 & \cdots & \sigma_v^2 \\ \sigma_v^2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_v^2 \\ \sigma_v^2 & \cdots & \sigma_v^2 & \sigma_e^2 + \sigma_v^2 \end{bmatrix} = \sigma_e^2 \mathbf{I}_T + \sigma_v^2 \mathbf{i}\mathbf{i}' \quad (33)$$

is a $T \times T$ matrix with $\sigma_e^2 + \sigma_v^2$ on the principal diagonal and σ_v^2 elsewhere, and $\mathbf{i}$ is a T -vector of ones.

Estimation of β using GLS requires Ω to be consistently specified at least to a factor, so a consistent estimate the ratio of σ_e^2 to σ_v^2 in order to obtain a consistent estimate of $\hat{\Omega}$. The feasible GLS parameter estimates are then given by

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^N X_i' \hat{\Omega}^{-1} X_i \right)^{-1} \left(\sum_{i=1}^N X_i' \hat{\Omega}^{-1} y_i \right) = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y \quad (34)$$

with variance

$$\text{var}(\hat{\beta}_{RE}) = \left(\sum_{i=1}^N X_i' \hat{\Omega}^{-1} X_i \right)^{-1} = (X' \hat{\Omega}^{-1} X)^{-1}, \quad (35)$$

where the parameter standard errors are given by the square root of the diagonal elements of (35).

A consistent estimate of σ_e^2 can be found by estimating the FE model (23) because there are NT residuals regardless of how N and T behave as $NT \rightarrow \infty$. Dividing the FE sum of squared errors SSE_{FE} from (23) by $NT - N - K$ provides an unbiased and consistent estimate of σ_e^2 since the error terms for this model are just the e_{it} . But the natural estimator of σ_v^2 , the sample variance of the elements of $\hat{\gamma}$ is not consistent unless $N \rightarrow \infty$.

If $NT \rightarrow \infty$ as $N \rightarrow \infty$, σ_v^2 can be estimated consistently. Start with the regression

$$P_D y = P_D X \beta + \text{residuals}, \quad (36)$$

where $P_D \equiv I - M_D$ and the projection matrix $M_D \equiv I - D(D'D)^{-1}D'$. The between-groups estimator is then

$$\hat{\beta}_{BG} = (X'P_DX)^{-1}X'P_D y, \quad (37)$$

which is based on only N observations since all variables are the same for every observation in each group. It only exists if $N > K$ because the matrix $X'P_DX$ can have at most rank N . It is called the between estimator because it uses only the variation among the group means.

If the assumptions of the random effects model are violated, the $\hat{\beta}_{BG}$ estimator, if it exists, is in general biased and inconsistent. To see this, unbiased and consistent parameter estimates require that the moment condition

$$E((P_DX)'_{it}(y_{it} - X_{it}\beta)) = 0. \quad (38)$$

$(P_DX)_{it}$ is the row labeled it of the $NT \times K$ matrix P_DX . Since $y_{it} - X_{it}\beta = v_i + e_{it}$ and e_{it} are independent of everything else in (38), this condition is equivalent to no correlation between the v_i and elements of the X matrix.

The variance of the error terms in the between group regression, (36), is $\sigma_v^2 + \sigma_e^2/T$. Therefore, if we estimate the regression with N observations, divide the resulting sum of squared errors by $N - K$, and then subtract σ_e^2/T , we obtain a consistent estimate of σ_v^2 . If this estimate is negative, the RE model should probably not be estimated, but Greene (2018), for example, provides detail on alternative methods for estimation the covariance components.

The feasible GLS estimator for the RE parameters is matrix-weighted averages of the within-groups (FE) and the between-groups estimators. Fuller and Battese (1973, 1974) show that the GLS estimator of the RE model can be obtained by using LS with the transformed data, i.e.,

$$(I - \lambda P_D)y = (I - \lambda P_D)X\beta + \text{residuals} \quad (39)$$

where the scalar

$$\lambda \equiv 1 - \left(\frac{T\sigma_v^2}{\sigma_e^2} + 1 \right)^{-1/2} = 1 - \left(\frac{\sigma_e^2}{T\sigma_v^2 + \sigma_e^2} \right)^{1/2} \quad (40)$$

and the consistent variance estimators developed above are used in (40). This provides a convenient way to compare the pooled LS, FE, and RE models. The RE GLS estimator is more efficient than the FE and pooled LS estimators as long as $\lambda \in (0, 1)$. If $\lambda = 0$, then the estimator is equivalent to pooled LS, and if $\lambda = 1$, then the FE (within-groups) estimator. Note that the LS estimator on the pooled data can be expressed as the weighted sum of the between and within estimators developed above, i.e.,

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y = (X'X)^{-1}(X'M_DY + X'P_DY) \\ &= (X'X)^{-1}X'(M_DX\hat{\beta}_{FE} + P_DX\hat{\beta}_{BG}).\end{aligned}\quad (41)$$

78.7 Balanced Versus Unbalanced Panel Data

The above developments above assume a balanced panel (each cross section has the same number of time series). The FE model requires no adjustment for unbalanced panels other than the appropriately defined D matrix (i.e., each column of D will no longer have the same number of 1's). However, estimation of the RE model does require more than straightforward revision to account for the unbalanced nature of the data. If the panel is unbalanced, (40) becomes

$$\lambda_i \equiv 1 - \left(\frac{T_i \sigma_v^2}{\sigma_e^2} + 1 \right)^{-1/2} \quad (42)$$

where T_i is the number of observations in the i^{th} cross-section. Naturally, the group means are based on the different numbers of observations in each cross section. LS is then used on this transformed data:

$$\begin{pmatrix} y_1 - \lambda_1 \bar{y}_1 \\ \vdots \\ y_N - \lambda_N \bar{y}_N \end{pmatrix} = \begin{pmatrix} X_1 - \lambda_1 \bar{X}_1 \\ \vdots \\ X_N - \lambda_N \bar{X}_N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \text{residuals} \quad (43)$$

where $\bar{y}_i$, a vector of size T_i , is the mean of the elements of y_i and $\bar{X}_i$, a $T_i \times k$ matrix whose elements are the mean of the corresponding column X_i . Equation (43) simplifies to (39) if the panel is balanced.

78.8 Durbin-Wu-Hausman (DWH) Tests

Durbin (1954), Wu (1973, 1974), and Hausman (1978) developed specification tests to compare parameters of alternative estimators. These types of tests are used for many purposes, e.g., to choose between alternative estimators of a model (e.g., FE versus RE here), and to test alternative theoretical models (Chermak and Patrick

2001). See, for example, Patrick (2021) for details and examples of alternative applications. Here we are primarily concerned with using a DWH test to choose between the FE and RE models.

The DWH test is based on vectors of contrasts of parameters and variances. In general, consider two estimators providing parameter vectors $\hat{\theta}$ and $\tilde{\theta}$ and testing

$$H_0 : \text{plim}(\hat{\theta} - \tilde{\theta}) = 0 \text{ vs. } H_A : \text{plim}(\hat{\theta} - \tilde{\theta}) \neq 0.$$

Under the null hypothesis, if the difference between the root- T consistent estimators is root- T consistent, then it converges in distribution with mean $\mathbf{0}$ and limit normal distribution

$$\sqrt{T}(\hat{\theta} - \tilde{\theta}) \xrightarrow{d} N[\mathbf{0}, \Omega(\hat{\theta} - \tilde{\theta})],$$

where $\Omega(\hat{\theta} - \tilde{\theta})$ is the contrast in variance-covariance matrices in the limiting distribution. The DWH test statistic is then given by

$$m = (\hat{\theta} - \tilde{\theta})' \left(\frac{\Omega(\hat{\theta} - \tilde{\theta})}{T} \right)^{-1} (\hat{\theta} - \tilde{\theta}) \stackrel{a}{\sim} \chi^2(q). \quad (44)$$

That is, m is asymptotically distributed $\chi^2(q)$ under the null with degrees of freedom q , the number of parameters (associated with potential endogenous variables) being compared across the two models considered. The null hypothesis is rejected if $m > \chi^2_{\alpha}(q)$, where the level of statistical significance is indicated by α .

For example, in the context of instrumental variables, if the original X variables are valid instruments (i.e., not correlated with the contemporaneous error), then the LS estimator is consistent and preferred to the IV estimator with other valid instruments. Since both would then have the same probability limit, both provide consistent parameter estimates so the preference between the estimators is then based on the efficiency of the parameter standard errors (see, e.g., Patrick 2021 for detail and an example). The comparison of FE and RE is analogous.

In general, to carry out the test, we need a consistent estimate of

$$\Omega(\tilde{\theta} - \hat{\theta}) = \text{var}[\hat{\theta}] + \text{var}[\tilde{\theta}] - 2\text{cov}[\hat{\theta}, \tilde{\theta}]. \quad (45)$$

The last term on the right-hand side of this expression is relatively computationally difficult. If the estimator of $\hat{\theta}$ from the null hypothesis model is efficient, then $\text{var}[\hat{\theta}] = \text{cov}[\hat{\theta}, \tilde{\theta}]$ and (45) simplifies to

$$\Omega(\tilde{\theta} - \hat{\theta}) = \text{var}[\tilde{\theta}] - \text{var}[\hat{\theta}]. \quad (46)$$

In estimating the FE and RE models, the effects are specified as either fixed or random as developed above. This implies that the effects are either specified as

X variables or error components. If effects are correlated with the X variables then FE is consistent while RE is not. The FE β parameters are estimated using dummy variables (the so-called LSDV estimator) included as explanatory variables or by group mean differencing the variables and applying LS as shown above. Random effects (RE): the ε_{it} covariance matrix is specified as functions of the variances of the u_i , v_i , and e_{it} , and FGLS, ML, or GMM is used for estimation. Joint hypothesis tests of the effects can be carried out with an F-test (for the FE model) or Wald, LR, or LM test (for the RE model). A Durbin-Wu-Hausman (DWH) type test can be used to choose between the FE and RE estimators, among other things, as will be discussed later in this chapter. Patrick (2021), for example, presents detail on various uses for these types of tests.

For now, we focus on using a DWH test to infer if there is correlation between the effects and the explanatory variables, X . By effects in the discussion in this section, I am referring to effects that are specified as part of the error structure. We may also specify mixed effects models as illustrated in the examples below.

The simplest form of DWH test to compute was developed in Hausman (1978) and can be used to choose between the FE and RE specifications of the model if RE is efficient. If applicable, a relatively easy way to compute the test for the comparing parameter estimates from the alternative estimators. Hausman's procedure for testing independence between the explanatory variables and residuals in a model using any asymptotically efficient estimator and a consistent but asymptotically inefficient estimator is much less computationally burdensome than the statistics developed in Wu (1973, 1974). Applied to compare the FE and RE estimators, if effects are not correlated with X , RE is consistent and efficient under the null hypothesis, and FE is consistent but inefficient. If effects are correlated with X , then FE is consistent and efficient and RE is inconsistent.

Although the DWH test statistic presented in (47) is in the context of the RE and FE panel estimators, this is otherwise a relatively general form of the test. The DWH test for FE versus RE hinges on the correlation of the effects with the common error in the model to be estimated. The test statistic for RE versus FE is given by

$$m = \left(\hat{\beta}_{RE} - \hat{\beta}_{FE} \right)' \left(\hat{\Omega}_{FE} - \hat{\Omega}_{RE} \right)^{-1} \left(\hat{\beta}_{RE} - \hat{\beta}_{FE} \right) \overset{A}{\sim} \chi_k^2, \quad (47)$$

where $\hat{\beta}_{RE}$ is a vector of RE parameter estimates, $\hat{\beta}_{FE}$ is a vector of FE parameter estimates, both of size k , and $\hat{\Omega}_{FE}$ and $\hat{\Omega}_{RE}$ are consistent estimates of the covariance matrices defined above for the FE and RE estimators. The DWH test statistic, m , then has k degrees of freedom under the null hypothesis H_0 : RE residuals, ε , are not correlated with X . Rejecting the null hypothesis implies that FE would be the preferred estimator, because the RE parameter estimates are inconsistent (at least one of the explanatory variables is correlated with the random effects in the error structure). If the null is not rejected, there is no preference between FE and RE in terms of consistency, but RE would be preferred to FE because RE produces relatively efficient standard errors. This is a result of the between information in RE estimation, FE discards this information. Note that we can also include in the test

the parameters associated with effects that are specified as explanatory variables in mixed effects models (see Example 2 below for illustration).

The efficiency of the estimator requirement in the null hypothesis above can be relaxed. For example, if there is reason to use robust standard errors then the estimator is not fully efficient, so the relatively easy to compute form of the DWH test above is not applicable. If the estimator in the null hypothesis is not efficient, then direct estimation using auxiliary regressions (Davidson and MacKinnon 1984, 1987, and 1989, and MacKinnon 1992) or the paired bootstrap (Cameron and Trivedi 2005) can be used to estimate the $\text{var}(\hat{\theta} - \tilde{\theta})$, in particular the last term on the right-hand side of (45).

78.9 Testing Strict Exogeneity

A DWH test, or robust DWH test, could also be used to test strict exogeneity by contrasting β parameters and variances estimated by FE versus first difference (FD). Since strict exogeneity holds in the FD specification, pooled LS estimation of the FD equations yields consistent parameter estimates. However, Wooldridge (2010) provides regression based tests of strict exogeneity that are relatively easy to carry out, which are briefly presented below.⁹

Wooldridge (2010) suggests two equivalent regression approaches to test strict exogeneity. One way, Wooldridge (2010) tests strict exogeneity is by estimating

$$y_{it} = X_{it}\beta + X_{i,t+1}\delta + D_i\gamma + \varepsilon_{it}, \quad (48)$$

and testing the null hypothesis $H_0 : \delta = 0$ versus the alternative $H_1 : \delta \neq 0$.

The same test can also be carried out by using LS to estimate the first-difference (FD) form of the model,

$$\Delta y_{it} = \Delta X_{it}\beta + X_{i,t+1}\delta + \varepsilon_{it}, \quad (49)$$

and testing the same null hypothesis as with (48).

Either of these strict exogeneity tests are straightforward to apply and cost little in effort and computational time, yet the benefits of avoiding parameter bias and inconsistency may be substantial. So there is much to gain and little to lose by applying a test of strict exogeneity.

78.10 Empirical Application

This section provides two examples to illustrate the application of the estimators and specification tests covered above in two very different modeling situations. The first is an economics example using experimental data, providing a unique comparison of

⁹Grieser and Hadlock (2019) use simulations to show that these tests have “substantial” power.

the estimators and concepts above. The second is a financial model for asset pricing, where data is not nearly so nicely behaved as in the first example.

78.10.1 Example 1: Comparison of Pooled LS, Pooled LS with Cluster Robust SE, FE, RE (GLS), and RE (ML) Estimators Using Experimental Data

This example provides a comparison of five panel estimators and their relative consistency and efficiency using experimental data from Chermak et al. (2021). Chermak et al. analyze market, survey, and experimental data to estimate a residential water demand model. Water utility managers are interested in household water demand estimates for planning and operational purposes, including pricing, revenue, and providing an adequate supply of water. The experimental data provides a sample of utility customers' water demand levels as the price of water changes. The experiment is repeated for 10 rounds, providing the time periods for the panel. The market data from the utility and survey of the participants in the experiment data provide fixed characteristics that may be used as explanatory variables in the estimated demand model. The prices the consumers in the experiment face are set independently of any reaction that consumers (participants) have to price, conditional on their fixed household infrastructure and characteristics. Since these explanatory variables are fixed (deterministic or nonstochastic) for the pricing experiment, they are contemporaneously and strictly exogenous. This provides the opportunity to compare a range of basic panel data estimators and tests.

The demand for water by household j during time t (round) estimate is

$$w_{jt} = f_{jt}(p, B, \Phi, \Omega) \\ = \beta_0 + \beta_1 p_t + \beta_2 B_j + \sum_{i=3}^{15} \beta_i (p_t x_{ij}) + \sum_{i=16}^{33} \beta_i x_{ij} + v_j + \varepsilon_{jt} \quad (50)$$

where β_i , $i = 0, \dots, 33$, are the estimated parameters; the price of water at time t is p_t ; B_j is the budget for household j , interaction terms between price and the time effects, between price and household size, and between price and location, x_{ij} , $i = 3, \dots, 15$, and other household characteristics, x_{ij} , $i = 16, \dots, 33$. These control for differences across household water conservation technology (XERIC indicates low water use landscaping, LFTLT1 low-flow toilets, LFDW low-flow dishwasher, and LFWM for low-flow washing machine), quadrants of the city ($x_j = SW, NW, SE$, with the NE as the base), household size (PEOPLE), and other characteristics specified in the estimation results reported in Table 1.

As discussed above, a key feature of this data is that all explanatory variables are deterministic and exogenous to the models. Nineteen of the explanatory variables are constant over time. Given this, the associated parameters and standard errors cannot be estimated with FE because these time constant variables are perfectly correlated with the cross-sectional fixed effects in the model. In Table 1, the FE column labels

the parameters to these specific constant explanatory variables “omitted.” The **bold** numerical values indicate parameter estimates where hypothesis test conclusions are conflicting across the different estimators (in terms of the indicated significance level). The pooled LS estimates provide the largest number of parameters that are indicated to be significant, but the LS standard errors are not efficient. Many Type I errors (rejecting the null hypothesis when it is true) would be made in this case. If one were to adopt the pooled LS model, one would draw many incorrect conclusions based on the indicated significance of the parameters, and be overly optimistic in any forecasts based on the standard error of the forecast.

With the exception of the parameters that could not be estimated with FE, the parameter estimates across all of the estimators are the same (for the decimal places shown). As we move across the estimators in Table 1 from left to right, the efficiency of the estimates increases (from pooled LS to pooled LS with cluster robust standard errors to FE to RE, and from RE estimated with GLS to the most efficient, RE estimated with ML). Based on efficiency of estimated standard errors, using pooled LS with cluster robust standard errors is clearly preferred to pooled LS. But the Type I error caveat discussed above remains, though not as severe. Neither of these models considers any cross-sectional effect and efficiency can be improved with other models.

All but one of the time effect interactions with price terms are highly significant individually (joint tests of these effect interactions are also all highly significant). Cross-sectional effects are also all highly significant, as indicated by the F-test for the FE model and Likelihood ratio (LR) and Lagrange multiplier (LM) tests for the RE ML and GLS estimates, respectively. Since all the parameter estimates are the same, it is clear that the random effects are not correlated with the explanatory variables in the model. There is no point in trying to calculate DWH tests for the common parameter estimates across the models. All of the estimators provide consistent estimates in this example, but the RE ML estimates are consistent and *fully* efficient, so they are naturally preferred in this case.

78.10.2 Example 2: FE and RE, Contemporaneous and Strict Exogeneity, DWH and Wooldridge Tests

Fama and MacBeth (1973) used a two-step estimation process in an early test of CAPM theory. Their first step was to use time series data to estimate separate regressions for each asset, obtaining an estimate of all β_i . Second, for each point in time, cross-sectional data was used to estimate the model

$$R_{it} - R_{ft} = \theta_0 + \theta_1 \beta_i + e_i \quad (51)$$

where $R_{it} - R_{ft} \equiv$ excess return on asset i at time t , β_i is the estimated beta, and θ_0 and θ_1 are the parameters to estimate. Fama and MacBeth estimated this second stage separately for each time period and calculated the average of the parameter estimates

Table 1 Comparing different panel data estimators (N > T)

Variable	Pooled LS	Pooled LS cluster robust SE	FE	RE (GLS)	RE (ML)	SE _{ML} -SE _{GLS}
PRICE	-0.26953725***	-0.26953725***	-0.26953725***	-0.26953725***	-0.26953725***	-0.000097
BUDGET	0.00077263***	0.00077263***	(omitted)	0.00077263***	0.00077263***	-0.000017
PEOPLE	1.5654565***	1.5654565***	(omitted)	1.5654565***	1.5654565***	-0.025583
SW	-4.4361875***	-4.4361875***	(omitted)	-4.4361875***	-4.4361875***	-0.097090
NW	-0.13281722	-0.13281722	(omitted)	-0.13281722	-0.13281722	-0.073780
SE	-1.6060862**	-1.6060862	(omitted)	-1.6060862	-1.6060862	-0.089073
OWN	2.6894308***	2.6894308	(omitted)	2.6894308	2.6894308	-0.127155
NNM	2.2718399***	2.2718399	(omitted)	2.2718399	2.2718399	-0.162117
LVENM	5.6425551***	5.6425551	(omitted)	5.6425551*	5.6425551*	-0.191083
LVEUS	1.6450165**	1.6450165	(omitted)	1.6450165	1.6450165	-0.157465
XERIC	-5.5348475***	-5.5348475***	(omitted)	-5.5348475***	-5.5348475***	-0.085007
SQFT	0.00104162***	0.00104162*	(omitted)	0.00104162**	0.00104162***	-0.000024
Swamp1	1.2605018***	1.2605018	(omitted)	1.2605018	1.2605018	-0.067471
LFTLT1	-2.2398105***	-2.2398105*	(omitted)	-2.2398105**	-2.2398105**	-0.066252
LFWM1	-3.2170521***	-3.2170521***	(omitted)	-3.2170521**	-3.2170521***	-0.072302
LFDW	1.4576449***	1.4576449	(omitted)	1.4576449	1.4576449	-0.084770
RN	-0.29299905	-0.29299905	(omitted)	-0.29299905	-0.29299905	-0.079858
RS	1.6571338***	1.6571338	(omitted)	1.6571338	1.6571338	-0.092438
RI	2.0633291***	2.0633291	(omitted)	2.0633291	2.0633291	-0.085557
SW#c.PRICE	0.15202784**	0.15202784***	0.15202784***	0.15202784***	0.15202784***	-0.000113
NW#c.PRICE	-0.02586791	-0.02586791	-0.02586791	-0.02586791	-0.02586791	-0.000090
SE#c.PRICE	0.0280335	0.0280335	0.0280335	0.0280335	0.0280335	-0.000110
c.PRICE#	-0.03222381	-0.03222381*	-0.03222381***	-0.03222381***	-0.03222381***	-0.000030
c.PEOPLE						
expptarci~n	-5.1854643***	-5.1854643***	(omitted)	-5.1854643***	-5.1854643***	-0.112140
RND#c.PRICE						

(continued)

Table 1 (continued)

Variable	Pooled LS	Pooled LS cluster robust SE	FE	RE (GLS)	RE (ML)	SE _{ML} -SE _{GLS}
1	-0.99544667**	-0.99544667***	-0.99544667***	-0.99544667***	-0.99544667***	-0.000590
2	-0.90281912***	-0.90281912***	-0.90281912***	-0.90281912***	-0.90281912***	-0.000415
3	-0.74207091***	-0.74207091***	-0.74207091***	-0.74207091***	-0.74207091***	-0.000290
4	-0.56640668***	-0.56640668***	-0.56640668***	-0.56640668***	-0.56640668***	-0.000201
5	-0.37584142***	-0.37584142***	-0.37584142***	-0.37584142***	-0.37584142***	-0.000139
6	-0.27644403***	-0.27644403***	-0.27644403***	-0.27644403***	-0.27644403***	-0.000109
7	-0.2035468***	-0.2035468***	-0.2035468***	-0.2035468***	-0.2035468***	-0.000086
8	-0.11615192**	-0.11615192***	-0.11615192***	-0.11615192***	-0.11615192***	-0.000070
9	-0.05338751	-0.05338751***	-0.05338751***	-0.05338751***	-0.05338751***	-0.000058
_cons	11.139752***	11.139752***	16.280120***	11.139752***	11.139752***	-0.218446
legend:	*	$p < 0.1$;	**	$p < 0.05$;	***	$p < 0.01$
testing						Testing
c.s. effects			F(177,1766) =106.3	LM(1) = 7269.8	LR(1) = 3118.9	c.s. effects
jointly testing X	F(33,1924) =37.4	F(33,1924) =37.4	F(14,1766) =75.9	Wald(33) =1165.1	LR(33) = 927.5	Jointly testing X

$T = 11$, $N = 178$, $NT = 1958$. This is a balanced panel. 19 of the X variables in (50) are constant through time for each cross section, so they are perfectly correlated with the effects in FE and can't be estimated.

to conduct hypothesis tests. However, a panel data approach could be used to carry out the second stage and directly test the Fama and MacBeth hypotheses.

Lee et al. (2019) point out that both cross section and time effects are of interest in financial analyses. To include time effects, the model in (51) is specified to include time dummies as additional explanatory variables. The corresponding panel data model with firm and time effects (a two-way panel data model) is specified as

$$R_{it} - R_{ft} = \theta_0 + \theta_1 \beta_i + v_i + u_t + e_{it} \quad (52)$$

where the cross-sectional effects v_i and time effects u_t could be specified as fixed, random, or mixed. Specification choice is examined below.

Data from Brooks (2014) on annual returns and estimated betas for 2500 UK firms over 11 years is used to estimate two-way FE and RE specifications of (51) and compare these estimates to those obtained using LS on the pooled data, as well as each other. This is an unbalanced panel data set. A DWH test is used to choose between these alternative specifications, i.e., to infer if effects are contemporaneously correlated with the explanatory variables. Wooldridge (2010) strict exogeneity tests are also performed.

Below are the estimation results (using STATA17) for

1. The pooled LS model
2. The FE model (two-way, firm and time effects)
 - (a) F-test of the effects
 - (b) Test of strict exogeneity.
3. Estimate the RE models (two-way, firm and time effects).
 - (a) Tests of the effects
4. DWH test to choose between FE and RE specifications

While cross-sectional effects are included in the model, parameters for cross-sectional effects are not presented below, but joint tests for cross-sectional effects are provided. The pooled LS estimation results are in Table 2. There are no significant parameter estimates indicated, the model lacks explanatory power. Estimates here may be biased and are most likely inefficient.

Table 2 Pooled LS

. reg return beta						
Source	SS	df	MS	Number of obs = 8,846		
				F(1, 8844) = 0.03		
Model	0.000084304	1	0.000084304	Prob > F = 0.8607		
Residual	24.2017921	8,844	0.002736521	R-squared = 0.0000		
				Adj R-squared = -0.0001		
Total	24.2018765	8,845	0.002736221	Root MSE = 0.05231		
return	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
beta	0.0004806	0.0027382	0.18	0.861	-0.004887	0.0058482
_cons	0.0018154	0.0030781	0.59	0.555	-0.0042184	0.0078492

FE estimation results, with firm and time effects, are presented in Table 3. Contrary to the pooled LS estimation results, the beta coefficient is now negative and highly significant. Jointly testing the time effects gives the test statistic $F(10,7106) = 63.14$, which implies the joint zero null hypothesis is rejected, so time effects are jointly highly significant. From the reported t-tests, all but two of the time effects are individually significant. Firm effects are also jointly significantly different than zero, as indicated by the F-test presented at the bottom of Table 3.

Using (48), the regression to test strict exogeneity is in Table 4. The coefficient on *lead1beta* is not significantly different than 0, so there is no indication of a violation of strict exogeneity for this model according to Wooldridge’s test. If we had found evidence of a departure from strict exogeneity, then we would try to model the source of endogeneity or eliminate it through model specification. For example, see Wintoki et al. (2012) for an example using a dynamic panel GMM estimator (in a notable exception in the literature, they acknowledge strict exogeneity). See Grieser and Hadlock (2019) for more on strict exogeneity tests and IV estimation in financial

Table 3 FE with firm and time effects

. xtreg return beta i.year, fe						
Fixed-effects (within) regression				Number of obs = 8,846		
Group variable: firm_id				Number of groups = 1,729		
R-squared:				Obs per group:		
Within = 0.0827				min = 1		
Between = 0.0018				avg = 5.1		
Overall = 0.0569				max = 11		
				F(11,7106) = 58.22		
corr(u_i, Xb) = -0.0346				Prob > F = 0.0000		
return	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
beta	-0.0117774	0.0039592	-2.97	0.003	-0.0195387	-0.0040162
year						
1997	-0.0043052	0.002146	-2.01	0.045	-0.008512	-0.0000984
1998	-0.0171009	0.0025746	-6.64	0.000	-0.0221479	-0.012054
1999	0.021552	0.0026082	8.26	0.000	0.0164391	0.0266649
2000	-0.0045447	0.0026897	-1.69	0.091	-0.0098173	0.0007278
2001	-0.0138006	0.0027694	-4.98	0.000	-0.0192293	-0.0083718
2002	-0.023654	0.0028292	-8.36	0.000	-0.0292	-0.018108
2003	0.0174809	0.002882	6.07	0.000	0.0118312	0.0231305
2004	-0.0064338	0.0029388	-2.19	0.029	-0.0121947	-0.0006728
2005	-0.0109381	0.0029809	-3.67	0.000	-0.0167815	-0.0050947
2006	0.0002718	0.0030602	0.09	0.929	-0.0057272	0.0062708
_cons	0.0189027	0.0047929	3.94	0.000	0.0095073	0.0282982
sigma_u	0.04170847					
sigma_e	0.04869689					
rho	0.42315824 (fraction of variance due to u_i)					
F test that all u_i = 0: F(1728, 7106) = 1.42				Prob > F = 0.0000		

Table 4 Test of strict exogeneity

. xtreg return beta lead1beta i.year, fe						
Fixed-effects (within) regression				Number of obs = 7,111		
Group variable: firm_id				Number of groups = 1,596		
R-squared:				Obs per group:		
Within = 0.1365				min = 1		
Between = 0.0223				avg = 4.5		
Overall = 0.1125				max = 10		
				F(11,5504) = 79.11		
corr(u_i, Xb) = -0.0498				Prob > F = 0.0000		
return	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
beta	-0.018281	0.0041527	-4.40	0.000	-0.026422	-0.01014
lead1beta	0.0022036	0.0044584	0.49	0.621	-0.0065366	0.0109438
year						
1997	-0.0078875	0.0027765	-2.84	0.005	-0.0133306	-0.0024445
1998	-0.0203879	0.0025383	-8.03	0.000	-0.0253641	-0.0154118
1999	0.0187852	0.0025938	7.24	0.000	0.0137004	0.02387
2000	-0.0093265	0.0026619	-3.50	0.000	-0.0145448	-0.0041081
2001	-0.0184515	0.0027057	-6.82	0.000	-0.0237558	-0.0131472
2002	-0.0268569	0.0027499	-9.77	0.000	-0.0322478	-0.021466
2003	0.0196439	0.002795	7.03	0.000	0.0141646	0.0251232
2004	-0.0052338	0.0028285	-1.85	0.064	-0.0107788	0.0003113
2005	-0.0067835	0.0028918	-2.35	0.019	-0.0124525	-0.0011144
_cons	0.0270216	0.0057012	4.74	0.000	0.015845	0.0381982
sigma_u	0.01857285					
sigma_e	0.04220725					
rho	0.16222269 (fraction of variance due to u_i)					
F test that all u_i=0: F(1595, 5504) = 0.59				Prob > F = 1.0000		

panel models. If valid instruments are available, contemporaneous correlation can be inferred by using IV estimation and a DWH test (see Patrick 2021 for more information and examples).

Next we consider the RE estimator and use a DWH test to choose between FE and RE. If the effects are not correlated with the explanatory variables X , then RE is consistent and more efficient than the FE. RE estimates using GLS are in Table 5.

Contrary to the FE results above, the beta coefficient is now positive but not significant at the usual levels. The Wald test with 11 degrees of freedom is to test the joint zero null for the beta and time effects, Wald(11) = 668.22, so these parameters are jointly highly significant. Table 6 provides a test of the cross-sectional random effect, LM(1) = 4.79 indicates that the random cross-sectional effect is highly significant. Table 7 contains a Wald test of the time fixed effects specified in the RE model, the Wald(10) = 664.98 implies that the fixed time effects are jointly highly significant.

Table 5 RE with firm and time effects

Random-effects GLS regression				Number of obs = 8,846		
Group variable: firm_id				Number of groups = 1,729		
R-squared:				Obs per group:		
Within = 0.0813				min = 1		
Between = 0.0015				avg = 5.1		
Overall = 0.0625				max = 11		
				Wald chi2(11) = 668.22		
corr(u_i, X) = 0 (assumed)				Prob > chi2 = 0.0000		
return	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
beta	−0.0052411	0.0031406	−1.67	0.095	−0.0113965	0.0009143
year						
1997	−0.0028082	0.0020655	−1.36	0.174	−0.0068564	0.0012401
1998	−0.0122569	0.0021679	−5.65	0.000	−0.0165059	−0.0080079
1999	0.0259504	0.0022224	11.68	0.000	0.0215946	0.0303062
2000	0.0008355	0.0023136	0.36	0.718	−0.0036991	0.0053702
2001	−0.008367	0.0023951	−3.49	0.000	−0.0130612	−0.0036728
2002	−0.0179014	0.0024561	−7.29	0.000	−0.0227153	−0.0130875
2003	0.0234545	0.0025092	9.35	0.000	0.0185367	0.0283724
2004	0.0001032	0.0025701	0.04	0.968	−0.0049341	0.0051404
2005	−0.004175	0.0026149	−1.60	0.110	−0.0093002	0.0009501
2006	0.0073301	0.0026971	2.72	0.007	0.0020438	0.0126163
_cons	0.0055398	0.0038647	1.43	0.152	−0.0020348	0.0131144
sigma_u	0.02852232					
sigma_e	0.04869689					
rho	0.25543037 (fraction of variance due to u_i)					

Table 6 LM test of cross-sectional random effect

. xttest0		
Breusch and Pagan Lagrangian multiplier test for random effects		
return[firm_id,t] = Xb + u[firm_id] + e[firm_id,t]		
Estimated results		
	Var	SD = sqrt(Var)
return	0.0027362	0.0523089
e	0.0023714	0.0486969
u	0.0008135	0.0285223
Test: Var(u) = 0		
chibar2(01) = 4.79		
Prob > chibar2 = 0.8143		

To summarize Example 2 results so far, FE results indicate that impact of beta is negative and significant. The fixed cross-sectional effects are significant, as are the fixed time effects. Testing strict exogeneity reveals no indication of a significant

Table 7 Wald test of time series fixed effects

. test (1997.year 1998.year 1999.year 2000.year 2001.year 2002.year 2003.year 2004.year 2005.year 2006.year)
(1) 1997.year = 0
(2) 1998.year = 0
(3) 1999.year = 0
(4) 2000.year = 0
(5) 2001.year = 0
(6) 2002.year = 0
(7) 2003.year = 0
(8) 2004.year = 0
(9) 2005.year = 0
(10) 2006.year = 0
chi2(10) = 664.98
Prob > chi2 = 0.0000

violation, so there is no indication of significant correlation between contemporaneous returns and future values of beta for the asset. The RE results indicate an estimate of the impact of beta that is positive but not significant at the usual levels. Joint tests of the time and cross-sectional effects indicate both are highly significant.

A DWH test is used to choose between the FE and RE specifications, i.e., to infer if effects are contemporaneously correlated with the explanatory variables. If effects are correlated with the explanatory variables, then RE will be inconsistent, so the FE model would be preferred (see the DWH discussion above). The DWH test statistic is presented in Table 8, the $\chi^2(11)$ distributed statistic is 38.90, so the null is rejected. From this we infer that the random effect is correlated with the common error, so RE is inconsistent in this case and we reject the RE model.

78.11 Conclusion

Advantages of panel data relative to a cross-sectional or time series model include increased precision in estimation through increased sample size, greater flexibility in identifying the DGP, amelioration of omitted variables problems, and relaxing the constant parameter assumption.

Statistical properties of panel data estimators depend on the specific model and specification of the cross-sectional unobserved effects. The LS estimator with panel data will typically overstate precision (underestimate the standard errors) because it does not control for likely correlation of regression errors over time for the cross section.

There is the possibility of consistent estimation of the FE model, which allows for unobserved heterogeneity that may be correlated with the explanatory variables. The most commonly used panel data estimator is FE, usually only considering variation in the cross section (or time series, these are one-way models) or both (two-way

Table 8 DWH for FE versus RE specification of the model

. hausman FE REgls				
	—— Coefficients ——			
	(b) FE	(B) REgls	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
beta	−0.0117774	−0.0052411	−0.0065363	0.0024109
year				
1997	−0.0043052	−0.0028082	−0.0014971	0.0005823
1998	−0.0171009	−0.0122569	−0.004844	0.0013888
1999	0.021552	0.0259504	−0.0043984	0.0013653
2000	−0.0045447	0.0008355	−0.0053803	0.0013717
2001	−0.0138006	−0.008367	−0.0054336	0.0013903
2002	−0.023654	−0.0179014	−0.0057526	0.0014042
2003	0.0174809	0.0234545	−0.0059737	0.0014178
2004	−0.0064338	0.0001032	−0.0065369	0.0014253
2005	−0.0109381	−0.004175	−0.0067631	0.001431
2006	0.0002718	0.0073301	−0.0070583	0.0014459
b = Consistent under H0 and Ha; obtained from xtreg .				
B = Inconsistent under Ha, efficient under H0; obtained from xtreg .				
Test of H0: Difference in coefficients not systematic				
$\text{chi2(11)} = (b-B)'[(V_b-V_B)^{-1}](b-B)$				
= 38.90				
Prob > chi2 = 0.0001				

models). Much of the literature treats the effects as nuisance (incidental) parameters that can be ignored since primary interest is generally in consistent estimation of the vector β . If the parameters associated with the fixed effects are of interest, they can be estimated by specifying the dummy variables and computing the within (FE) estimator by least squares dummy variable estimation.

Some explanatory variables may have no within group variation because they are only observed at the group level. Such variables will exhibit perfect multicollinearity with group dummies in the FE model and, as a result, the associated parameters will not be identified. However, they may be identified by the RE model since RE take into account between-group variation. That is, FE parameters of time-invariant regressors are not identified in the within model (if $x_{it} = x_i$, then $\bar{x}_i = x_i$ and $x_{it} - \bar{x}_i = 0$). FE with group effects cannot be used in this case because it violates the full rank assumption.

To estimate the effect of time-invariant regressors, we could potentially use pooled LS or random effects (RE) estimators since they allow the time invariant effects to be estimated. The effects are captured in the error for both pooled LS and RE estimators and both will produce inconsistent parameter estimates if the effects are correlated with the explanatory variables. Other FE specifications may still be estimable (e.g., time invariant regressors and time rather than group effects).

Error component (see Chen, et al. (2021), Lee and Jen (2021), Kim (Chapter 82), and Chen, et al. (Chap. 103, “Alternative Errors-in-Variables Models and Their

Applications in Finance Research”)) and clustering models are also relevant for cross section data where subsets of observations are natural groups. For example, consider a cross section of firms from different countries, it may be that the error terms for firms within each country are correlated.

ML and GMM are also commonly used estimators for panel data models. The algebra of the linear panel data estimators will often not generalize to nonlinear panel estimators (see Greene (2018)).

The first-difference (FD) estimator may also be useful with panel data. In short panels, $N > T$, it measures the relationship between individual specific one period changes in the regressor and individual-specific one-period changes in the dependent variable. Consistency requires $E[e_{it} - e_{i, t-1} | X_{it} - X_{i, t-1}] = 0$, less stringent than strong exogeneity (the within estimator), but a stronger condition than $E[e_{it} | X_{it}] = 0$.

There may be endogeneity issues in panel data, so contemporaneous and strict exogeneity should be tested. In theory, endogeneity issues can be corrected using IV methods for each cross section, but valid instruments are necessary. Hausman-Taylor (1981b) and Amemiya and MaCurdy (1986) estimators allow instrumental variables in panel models. On validity of instruments, see Hahn and Hausman (2002) and Hausman et al. (2005).

78.12 Cross-References

► [Alternative Errors-in-Variables Models and Their Applications in Finance Research](#)

References

- Amemiya, T., and T.E. MaCurdy. 1986. Instrumental-variable estimation of an error-components model. *Econometrica* 54: 869–881.
- Baltagi, Badi H. 2021. *Econometric analysis of panel data*. 6th ed. Cham: Springer.
- Brooks, C. 2014. *Introductory econometrics for finance*. 3rd ed. Cambridge: Cambridge University Press.
- Cameron, A.C., and P.K. Trivedi. 2005. *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press.
- Chamberlain, Gary. 1984. Panel Data. In *Handbook of econometrics*, ed. Z. Griliches and M.D. Intriligator, vol. II. Amsterdam: Elsevier Science Publishers BV.
- Chang, Hui-shyong, and Cheng F. Lee. 1977. Using pooled time-series and cross section data to test the firm and time effects in financial analyses. *Journal of Financial and Quantitative Analysis* 12 (3): 457–471.
- Chen, Hong-Yi, Alice C. Lee, and Cheng Few Lee. 2021. Alternative methods to deal with measurement error (Ch. 37). In *Handbook of financial econometrics, mathematics, statistics, and machine learning*, ed. C.F. Lee and J.C. Lee, vol. II, 1439–1484. New Jersey: World Scientific.
- Chermak, J.M., and R.H. Patrick. 2001. A microeconomic test of the theory of exhaustible resources. *Journal of Environmental Economics and Management* 42 (1): 82–103.
- Chermak, J.M., K. Krause, D. Brookshire and R.H. Patrick. 2021. Residential water consumer response to Price: A field experiment inquiry. Working Paper, University of New Mexico.

- Davidson, R., and J.G. MacKinnon. 1984. Model specification tests based on artificial linear regressions. *International Economic Review* 25: 485–502.
- . 1987. Implicit alternatives and the local power of test statistics. *Econometrica* 55 (6): 1305–1329.
- Davidson, R., and J. MacKinnon. 1989. Testing for consistency using artificial regressions. *Econometric Theory* 5 (3): 363–384.
- Davidson, R., and J.G. MacKinnon. 2004. *Econometric theory and methods*. New York: Oxford University Press.
- Durbin, J. 1954. Errors in variables. *Review of the International Statistical Institute* 22: 23–32.
- Engle, R.F., D.F. Hendry, and J.-F. Richard. 1983. Endogeneity. *Econometrica* 51 (2): 277–304.
- Fama, E.F., and J.D. MacBeth. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 81 (3): 607–636.
- Frisch, R., and F. Waugh. 1933. Partial time regressions as compared with individual trends. *Econometrica* 1: 387–401.
- Fuller, W.A., and G.E. Battese. 1973. Transformations for linear models with nested error structure. *Journal of the American Statistical Association* 68: 626–652.
- . 1974. Estimation of linear models with crossed-error structure. *Journal of Econometrics* 2: 67–78.
- Giglio, Stefano, and Dacheng Xiu. 2021. Asset pricing with omitted variables. *Journal of Political Economy* 129(7):1947–1990.
- Greene, W.H. 2018. *Econometric analysis*. 8th ed. Upper Saddle River: Prentice Hall.
- Grieser, W.D., and C.J. Hadlock. 2019. Panel-data estimation in finance: Testable assumptions and parameter (in)consistency. *Journal of Financial and Quantitative Analysis* 54 (1): 1–29.
- Hahn, J., and J. Hausman. 2002. A New Specification Test for the Validity of Instrumental Variables. *Econometrica* 70(1): 163–189.
- Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1276.
- Hausman, J.A., and W.E. Taylor. 1981a. A generalized specification test. *Economics Letters* 8: 239–245.
- . 1981b. Panel data and unobservable individual effects. *Econometrica* 49: 1377–1398.
- Hausman, J., J.H. Stock, and M. Yogo. 2005. Asymptotic properties of the Hahn-Hausman test for weak instruments. *Economics Letters* 89: 333–342.
- Hsiao, Cheng. 2014. *Analysis of panel data*. 3rd ed. Cambridge: Cambridge University Press.
- Lee, C.F., and F.C. Jen. 2021. Effects of measurement errors on systematic risk and performance measure of a portfolio (Ch. 62). In *Handbook of financial econometrics, mathematics, statistics, and machine learning*, ed. C.F. Lee and J.C. Lee, vol. II, 2251–2264. New Jersey: World Scientific.
- Lee, C.F., Hong-Yi Chen, and John Lee. 2019. *Financial econometrics, mathematics, and statistics: Theory, method and application*. New York: Springer.
- Lovell, M. 1963. Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association* 58: 993–1010.
- MacKinnon, James G. 1992. Model specification tests and artificial regressions. *Journal of Economic Literature* XXX: 102–146.
- Patrick, Robert H. 2021. Durbin-Wu-Hausman specification tests (Ch. 28). In *Handbook of financial econometrics, mathematics, statistics, and machine learning*, ed. C.F. Lee and J.C. Lee, vol. II, 1075–1108. New Jersey: World Scientific.
- Pesaran, M. Hashem. 2015. *Time series and panel data econometrics*. Oxford: Oxford University Press.
- Wintoki, M.B., J.S. Linck, and J.M. Netter. 2012. Endogeneity and the dynamics of internal corporate governance. *Journal of Financial Economics* 105 (3): 581–606.
- Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. 2nd ed. Cambridge, MA: MIT Press.
- Wu, D. 1973. Alternative tests of Independence between stochastic regressors and disturbances. *Econometrica* 41: 733–775.
- . 1974. Alternative tests of independence between stochastic regressors and disturbances: Finite sample results. *Econometrica* 42: 529–546.



Accruals and the Asymmetric Timeliness of Earnings: A Decomposition Analysis

79

Wenhsin Hsu

Abstract

Since Basu (*Journal of Accounting and Economics*, 24, 3–37, 1997), the difference between the response of earnings to good news (as proxied by positive stock return) and its response to bad news (as proxied by negative stock return) has been referred to as evidence of asymmetric timeliness in the response of earnings to good and bad news and the difference has been termed a measure of asymmetric timeliness. This measure of asymmetric timeliness has been extensively used in assessing accounting conservatism and its economic consequences. In this study, I seek to contribute to the debate on the sources of asymmetric timeliness by testing for its existence across a range of components of earnings. When clean surplus earnings are decomposed into cash flow and accrual components, I find that asymmetric timeliness in operating cash flow is larger than that in accruals. Moreover, using articulated data, the decomposition across line items also finds little evidence with regard to the application of three prominent standards (i.e., the LCM rule, the impairment tests, and the bad debt provision) that incorporate asymmetric verification criteria. These results call into question whether asymmetric timeliness of earnings is a consequence of accounting conservatism.

Keywords

Conservatism · Asymmetric timeliness · Basu (1997) measure · Decomposition of clean surplus earnings

W. Hsu (✉)

Department of Accounting, National Taiwan University, Taipei, Taiwan

e-mail: audreyhsu@ntu.edu.tw

79.1 Introduction

Since the seminal paper Basu (1997), the difference between the timeliness of earnings in recognizing good and bad news has been characterized as an indicator of earnings conservatism and the informativeness of accounting numbers. The literature (Basu 1997; Pope and Walker 1999) has documented that accounting earnings generally recognize bad news (proxied by negative returns) more quickly than good news (proxied by positive returns) and has extensively employed this measure, “asymmetric timeliness,” over different governance systems and accounting regimes to compare earnings conservatism and its economic consequences (Ball et al. 2000, 2003; Pope and Walker 1999). Although most empirical evidence supports the role of conservatism in causing asymmetric timeliness, the literature still has not provided conclusive evidence on the sources of asymmetric timeliness of earnings. In particular, Dietrich et al. (2007) have raised concern about the econometric appropriateness of using reverse regression to examine earnings conservatism. They argue that the observed asymmetric timeliness of earnings in Basu (1997) is an econometric effect, not a result of conservatism, as the effect can also be observed in their simulated data where accounting conservatism is by construction absent.

In this chapter, I shed further light on the sources of asymmetric timeliness by providing a comprehensive picture of the pattern of asymmetric timeliness across earning components. Specifically, I seek evidence as to whether or not the incidence of asymmetric timeliness across earnings components is consistent with the theory that asymmetric timeliness is due to accounting conservatism. In order to do this, I decompose earnings across two dimensions: into cash flow and accrual components, and into line items.

If asymmetric timeliness is the direct result of earnings conservatism, the asymmetric timeliness should only be associated with accrual components rather than cash flow components. Therefore, a decomposition analysis of accrual and cash flow components can determine the sources of asymmetric timeliness.

Furthermore, although conservatism is considered a pervasive feature of financial reporting, it is likely to be observed in certain line items more than others. If asymmetric timeliness of earnings is a measure for earnings conservatism, asymmetric timeliness should predominately take place in the components, to which a standard with asymmetric requirement for gains and losses is directly related. I focus on three main standards, LCM rule, impairment tests and bad debt (warranty) provision, and the corresponding components.

One innovative feature of the study is that the asymmetric timeliness in aggregated earnings can be disaggregated into asymmetric timeliness in components. Specifically, in the Basu (1997) model where the “aggregated earnings” are regressed on the news, the proxy of which is stock return, the “timeliness” coefficient for news is determined by the covariance of “aggregated earnings” and stock return, divided by the variance of stock return. If the dependent variable (aggregated earnings) is decomposed into a series of earnings components, given that the denominator (the variance of stock returns) is constant, the “timeliness” coefficients

in each of the subsumed components are additive. Consequently, the asymmetric timeliness of clean surplus earnings is exactly identical to the sum of asymmetric timeliness in the subsumed components. This feature is helpful in depicting the sources of asymmetric timeliness of earnings.

With a unique dataset in which clean surplus earnings are articulated with net dividends and consecutive changes in book value, I seek evidence on earnings components subsumed in clean surplus earnings. The results show that asymmetric timeliness in operating cash flow is larger than that in accruals. The decomposition across line items also finds little evidence with regard to the application of three prominent standards that incorporate asymmetric verification criteria. The LCM rule might be expected to induce accounting conservatism, but no asymmetric timeliness is observed in the component of “change in inventory”; impairment tests might also be expected to induce accounting conservatism but little asymmetric timeliness is observed in special items; the rule of bad debt or warranty provision might be expected to induce accounting conservatism in working capital accruals due to the required recognition of a provision for future liability but not future benefits, but little asymmetric timeliness is observed in “working capital accruals.” These results call into question whether asymmetric timeliness of earnings is a consequence of conservatism.

The chapter is organized as follows. Section 79.2 discusses the theoretical background and hypothesis development. Section 79.3 describes research design, and Sect. 79.4 describes sample selection. Empirical evidence is presented in section 79.5. Section 79.6 presents alternative explanation and some additional tests, and Sect. 79.7 summarizes my finding and their implications.

79.2 Theoretical Background and Hypothesis Developments

79.2.1 Earnings Conservatism in the UK

One main purpose of reporting income is to make management accountable for its efforts on behalf of ownership. Moral hazard¹ is a type of information asymmetry that exists between managers and shareholders as ownership tends to be separated from control in most big entities. Because shareholders (principal) are unable to observe managers’ (agent) behaviors, managers tend to maximize their own benefits at the expense of the shareholders by taking advantage of their position of better information.

Watts (2003) proposes that accounting conservatism is a means of addressing moral hazard problems by constraining managers’ opportunistic behaviors in

¹Moral hazard refers to the tendency of the agent who is imperfectly monitored by the principal to engage in dishonest or otherwise undesirable activities.

reporting accounting measures. This is generally operated through the exercise of higher caution in the recognition of gains than losses. Such an asymmetric verification requirement can curb the enthusiasm of managers who want to maximize the earnings in order to distribute some net asset value to themselves when earnings is used in the compensation contracts.² Unless another efficient technology can be found to address the hazard problems between managers and shareholders, conservatism will continuously serve as a fundamental convention in accounting standards.

Specifically, earnings conservatism has been defined as accountants' tendency to require a higher verification for recognizing good news than bad news³ (Basu 1997; Watts 2003). This can induce asymmetry in the timeliness with which economic news is reflected in reported earnings: accounting earnings tend to recognize concurrent bad news as observed in negative market returns more quickly than good news as observed in positive returns. Basu (1997) terms the excess of the sensitivity of earnings to bad news over good news as "asymmetric timeliness of earnings." A firm is willing to supply the level of asymmetric timeliness of earnings to the extent that the benefits outweigh the costs of lower earnings (Hsu et al. 2017; Zhong and Li 2017; Burke et al. 2020).

Before highlighting the limitation of asymmetric timeliness in estimating earnings conservatism, I summarize some of the standards that incorporate the asymmetric verification requirement for recognizing gains and losses in UK GAAP:

1. *The Lower of Cost or Market Value Rule*

The lower of cost or market value is traditionally applied to stock valuation. SSAP 9 states that,

The amount at which stocks are stated in periodic financial statement should be the total of the lower of cost and net realisable value of the separate item of stock or of groups of similar items. (Para. 26)

While stock values are written down to net realizable value on an item-by-item basis or by reference to groups of similar items, they are not allowed to be written up if the net realizable value is higher than the carrying value. Once an asset is written down, the written-down value becomes the new cost, which cannot be written up again. The recognition of the written-down loss can be reflected in the decrease of inventory.

²Conservatism can also reduce the possibility that management make a liquidating dividend to shareholders at the expense of debt holders by manipulating earnings upwards. Other contracting explanations such as tax, litigation can refer to Watts (2003).

³There are other aspects of earnings conservatism. For example, the historical cost convention does not recognize the excess expected NPV (economic rent) at the time of acquisition; the choice of income-deferring accounting methods (i.e., depreciation or inventory).

2. *The Impairment Test for Tangible Assets, Intangible Assets, and Goodwill*

Another commonly illustrated example for the asymmetric recognition of gains and losses is the impairment test for assets. The ASB has issued a standard, FRS 11⁴ to account for the impairment of fixed assets and goodwill. When the book value of the assets exceeds the recoverable amount,⁵ capital assets should be written down to net recoverable value, but not vice versa if the recoverable amount exceeds the carrying value. The reduction of their carrying values is recorded as an “exceptional item” in the profit and loss account (Para. 67–73).

Nevertheless, one unique feature in UK GAAP is that the revaluation upwards⁶ is allowed for certain fixed assets, particularly the investment property. The revaluation surplus is credited to revaluation reserve as a dirty surplus flow rather than be recorded in the profit and loss account (Para. 65).

3. *Contingent Assets and Liabilities*

Under FRS 12, provisions, contingent liabilities, and contingent assets, companies are not allowed to recognize a contingent asset under any circumstances (Para. 31), but contingent liabilities should be recognized if a present obligation that arises from past events is probable and the amount of the obligation can be measured reliably (Para. 2). Likewise, the recognition of a provision for future economic benefits cannot in general be recognized, but a provision for future liability should be recognized if a reliable estimate can be made for the highly probable obligation. Typical examples are the bad debt provision that is reflected in the decrease of accounts receivable, the warranty liability that is recognized in the increase of accounts payable or the recognition of restructuring cost.

To sum up, it can be seen that the conservative practice with regards to asymmetric verification for gains and losses is an important feature of the UK accounting standards. An early recognition of losses in response to economic news mainly takes the form of asset write-offs or write-downs, and the establishment of loss provisions. The LCM rule applies to “changes in inventory” of current accruals; the impairment test applies to “special items”; and bad debt or warranty provisions apply to “changes in working capital.”

⁴The equivalent international standard is IAS 36 impairment of assets, issued in 1998. The FRS treats intangible assets in much the same way as goodwill while the IAS aligns their treatment to that of tangible assets.

⁵The recoverable amount is determined by the higher of its net realisable value and its value in use. Value in use is to be calculated by discounting the expected future cash flows at an estimate of the rate the market would expect on an equally risky investment.

⁶The Company Act 1985 permits either historical cost accounting or the use of an alternative basis of valuation. The issuance of FRS 15 in 1999 has removed considerable flexibility in the previous revaluation practice. FRS 15 requires companies to revalue the assets on a regular basis and does not permit revaluation of individual assets from same class.

79.2.2 Asymmetric Timeliness of Earnings in Earnings Constructs

79.2.2.1 Limitations of Asymmetric Timeliness

Following Basu (1997), asymmetric timeliness of earnings has been widely used as a measure of total conservatism. Although a great number of studies have exclusively employed the measure to assess the full extent of conservatism over countries and accounting regimes (Ball et al. 2000; Giner and Rees 2001; Pope and Walker 1999), recent studies have started to find that some limitations exist in the measure of “asymmetric timeliness” in estimating the full extent of conservatism. For example, Givoly et al. (2007) have documented that asymmetric timeliness can only capture one dimension of conservatism, and the asymmetric timeliness is on average inversely related to other dimensions of accounting conservatism. In other words, the asymmetric timeliness has difficulty in reflecting some conservatism policies regarding the choice of income-deferring accounting methods (such as straight line depreciation and accelerated depreciation) when the policies are not referenced to concurrent market news. But these conservative policies can inversely affect the asymmetric timeliness effect. This argument is consistent with Pope and Walker (2003) that asymmetric timeliness is a conditional measure that is negatively related to the ex ante conservatism (or balance sheet conservatism). This is because firms with higher capitalization of assets at the start of an accounting period tend to display stronger asymmetric timeliness than those with lower capitalization.

Furthermore, there is also some suspicion with regards to the econometric appropriateness of using reverse regression to examine accounting conservatism. Dietrich et al. (2007) have argued that the asymmetric timeliness of earnings is not a pure conservatism product, but an econometric effect. They argue that the model is mis-specified because the model relies on certain assumptions regarding the earnings-returns relation, and the model is only interpretable when returns cause earnings and not vice versa. Specifically, they point out that the Basu (1997) regression suffers from a sample variance ratio bias arising from the reverse regression when earnings are regressed on returns, and a sample truncation bias induced by splitting the sample by an endogenous variable (returns). Dietrich et al. (2007) argue that the observed asymmetric timeliness in Basu (1997) is completely explained by these econometric problems. To support this argument, Dietrich et al. (2007) also provide evidence that the asymmetric timeliness can be observed in data where, by construction, accounting conservatism is absent.

79.2.3 Hypothesis Developments

Since the literature does not provide conclusive evidence with regard to the sources of asymmetric timeliness of earnings, this study attempts to determine the asymmetric timeliness in all components of earnings. This is crucial in providing further evidence as to whether asymmetric timeliness is purely a consequence of earnings

conservatism. The decomposition analysis for the asymmetric timeliness of earnings is across two dimensions:

1. Cash Flow Versus Accruals

First, I decompose clean surplus earnings into cash flow and accrual components. Asymmetric timeliness of earnings has been hypothesized to be the product of earnings conservatism and the accrual process (Basu 1997). Through the application of earnings conservatism, the accrual process requires higher verification requirement for favorable news than unfavorable news and is expected to generate a differential speed for recognizing good and bad news. If asymmetric timeliness is the direct result of earnings conservatism, the asymmetric timeliness should be associated with accrual components and should not be associated with cash flow components. Therefore I expect that, if it is only earnings conservatism that determines asymmetric timeliness of earnings, asymmetric timeliness is only related to accrual components.

H1a: *Ceteris paribus*, asymmetric timeliness of earnings is completely driven by asymmetric timeliness in accrual components, and no asymmetric timeliness can be found in operating cash flow components.

Rejection of the above hypothesis implies that asymmetric timeliness of earnings is not a pure consequence of earnings conservatism.

2. Line Items

Second, I am also interested in earnings components across line items. It has been suggested that aggregate earnings should be decomposed into line items for a better understanding of the role of earnings in performance measurement or valuation as each component provides information that is complementary to information provided by the aggregate earnings (Lipe 1986). Therefore, earnings conservatism as observed in the asymmetric timeliness is better understood by reference to earnings components than by reference to aggregate earnings above.

Moreover, although conservatism is considered a pervasive feature of financial reporting, the “asymmetric timeliness” cannot capture all types of conservative policies (Givoly et al. 2007). Basu (1997, p. 16) reckons that only the accruals with regards to write-offs are more likely to reflect asymmetric timeliness type of conservatism and are reflected in extraordinary items. In other words, the prominent instances that asymmetric timeliness is expected to capture are the standards with asymmetric requirements for unrealized gains and losses, such as lower of cost or market value for inventories (LCM) or the impairment tests for fixed assets. This suggests that “asymmetric timeliness” is likely to be observed in earnings components where there is a specific standard with asymmetric verification requirements. Consequently, I predict that, if asymmetric timeliness of earnings is representative of earnings conservatism, asymmetric timeliness should predominately take place in the components, to which a standard with asymmetric requirement for gains and losses is directly related. I will focus on three main standards, LCM rule, impairment

tests and bad debt (warranty) provision, and the corresponding components: “working capital accruals,” and “exceptional items,”⁷ etc.

H1b: *Ceteris paribus*, the asymmetric timeliness of earnings components is stronger in the component where a standard exists which explicitly requires asymmetric recognition of gains and losses (i.e., working capital accruals, special items) than other components.

3. Dirty Surplus Flows

Finally, as I seek evidence on earnings components subsumed in clean surplus earnings, my analysis also covers the dirty surplus flows. Dirty surplus flows refer to gains and losses that bypass the profit and loss account. Under the clean surplus relationship, accounting earnings should include all changes in the book value of shareholders' equity during a period except the net contributions by owners. However, UK GAAP has permitted a wide use of dirty surplus flows (violation of clean surplus accounting) which allow that some of the gains and losses are not reflected in the accounting earnings, but are recognized directly in shareholders' equity. The typical example includes “prior-year adjustments,” “currency translation differences,” “goodwill write-offs,” and “revaluations,”⁸ and such treatment has created some concerns in practice.

The main concern⁹ for dirty surplus flows is that the violation of clean surplus accounting is inconsistent with accounting-based valuation models and is more subject to managers' manipulation because dirty surplus flows are more likely to be overlooked by financial users (Robinson 1991; Johnson et al. 1995). Although bad news is more likely to be concealed in dirty surplus flows than good news,

⁷Exceptional are material items which derive from events or transactions that fall within the ordinary activities of the reporting entity and which individually or, if of a similar type, in aggregate, need to be disclosed by virtue of their size or incidence if the financial statements are to give a true and fair view (FRS 3, Para.5). FRS 3 requires the “gains and losses from sales or termination of continued and discontinued operations,” “gains and losses from sales of fixed assets,” and “reorganisation” be separately disclosed as special items after operating profit.

⁸Among all dirty surplus flows, goodwill and revaluation reserve deserve some notice as they constitute the major part of dirty surplus flows. The goodwill treatment that requires the written-off against shareholders' equity in acquisition accounting (SSAP 22: *Accounting for Goodwill*) has long been criticised on the grounds that it overstates earnings and produce other mismeasurement problems. As such, FRS 10 (*Goodwill and Intangible assets*) eliminated dirty surplus treatment of goodwill write-offs in 1998. FRS 10 requires purchased goodwill and intangible assets to be capitalized and amortized through the profit and loss account over their useful economic lives. All goodwill previously eliminated against reserves has been reinstated as an asset on the balance sheet by way of a prior year adjustment.

⁹On the contrary, proponents of dirty surplus accounting argue that reported earnings should abstract from transitory effects to reflect current operating performance, enhancing the predictive ability for valuation purposes (Black 1993) because the inclusion of some unusual items in reported earnings is likely to cause misleading inferences, and managers are in a better position to identify those unusual items.

due to the obscure position in financial statements, I do not expect to observe asymmetric timeliness in dirty surplus flows. One reason is that current empirical evidence finds little evidence that dirty surplus flows are value relevant: the UK evidence by O'Hanlon and Pope (1999) and the USA evidence by Dhaliwal et al. (1999). Isidro et al. (2004) also show that dirty surplus flows except "goodwill flows" do not create bias in value estimates with respect to clean surplus-based value estimates.

Moreover, in 1992, FRS 3 has required the disclosure of dirty surplus flows in the "statement of total recognised gains and losses" and "reconciliation of movement in shareholders' funds" as a supplement to the "profit and loss account." This can create more transparency in dirty surplus flows, and reduce the tendency that bad news is hidden in dirty surplus flows. Therefore, I predict that, with the data from 1992 to 2000, asymmetric timeliness is not pronounced in dirty surplus flows.

H1c: *Ceteris paribus*, asymmetric timeliness cannot be observed in dirty surplus flows.

79.3 Research Design

In this chapter, I employ a decomposition analysis for asymmetric timeliness of earnings. Following Basu (1997), this study employs the asymmetric timeliness model, the intercept and coefficient of which varies with the sign of market returns. I will first introduce two decompositions of clean surplus earnings in Sect. 79.3.1. Section 79.3.2 describes a feature with regard to the equivalence between asymmetric timeliness in clean surplus earnings and the sum of asymmetric timeliness in each component.

79.3.1 Earnings Components

As the decomposition analysis of clean surplus earnings in this study is operated through two dimensions, I decompose clean surplus earnings across line items and across accrual and cash flow components. Exhibit 1 summaries all the disaggregated components in the study: the vertical dimension is related to decomposition across line items, and the horizontal dimension is for the decomposition across accrual and cash flow components. The line items are displayed in the context of the "profit and loss account," the "statement of total recognised gains and losses," and "reconciliation of movements in shareholders' funds" under UK GAAP (In Appendix 1, I extract financial statements of Boots plc to illustrate the format of three financial statements). Thus, the line items start mainly from operating profits, special items, tax, dirty surplus flows to the aggregated clean surplus earnings.

On the other hand, as accounting earnings is composed of accrual and cash flow components, the line item in each row can be decomposed into its accrual and cash

Exhibit 1 Earnings components of clean surplus earnings

Earnings	=	Cash flow	+	Accruals
Operating profit (OP_t)	=	Reported operating cash flows ($ROCF_t$)	+	Operating accruals (OA_t)
				[=working capital ($WCA_t = \Delta AR_t + \Delta INV_t - \Delta AP_t$) + Long-term operating accruals ($LOA_t = DEP_t + OOA_t$)] ^a
+ Special items ($Special_t$)	=	-	+	$Special_t$
[=gains (losses) from sales of fixed assets (Spe_FA_t) + gains(losses) from termination of operation ($Spec_ter_t$) + other special (Spe_oth_t)]				
+ Net interest expense (INT_t)	=	INT_t	+	-
+ Other nonoperating incomes (ONO_t)	=	-	+	ONO_t
= Pretax profits ($Pretax_t$)	=	$ROCF_t + INT_t$	+	$OA_t + Special_t + ONO_t$
+ Taxation (Tax_t)	=	Tax_t	+	-
= Net profit before extra and minority interest (EAR_t)	=	$ROCF_t + INT_t + Tax_t$	+	$OA_t + Special_t + ONO_t$
+ Extraordinary items ($Extra_t$)	=	-	+	$Extra_t$
+ Minority interest and preference dividend ($Minpref_t$)	=	$Minpref_t$	+	-
= Dirty surplus earnings (DSE_t)	=	$ROCF_t + INT_t + Tax_t + Minpref_t$	+	$OA_t + Special_t + ONO_t + Extra_t$
+ Dirty surplus flows (DSF_t)	=	-	+	DSF_t
[=prior year adjustment (PYA_t) + foreign exchange translation (FX_t) + revaluation (REV_t) + goodwill write-offs (GW_t) + other flows (OTH_t)]				
= Clean surplus earnings (CSE_t)	=	Total cash flows (TCF_t) [= $ROCF_t + INT_t + Tax_t + Minpref_t$]	+	Total accruals ($TACC_t$) [= $OA_t + Special_t + ONO_t + Extra_t + DSF_t$]

^a ΔAR_t , change in accounts receivables; ΔINV_t , change in inventory; ΔAP_t , change in accounts payable; DEP_t , depreciation; OOA_t , other operating accruals

flow component. Following the decomposition of each line item into corresponding accrual and cash flow components, those accrual components can sum up to total accruals as shown in the bottom line and all cash flow components can sum up to the aggregated cash flows. These two aggregated figures then add up to clean surplus earnings. Consequently, clean surplus earnings can be decomposed across accrual and cash flow components.

A. Decomposition Across Line Items

Specifically, the decomposition of clean surplus earnings across line items is expressed as:

$$\begin{aligned}
 CSE_t &= \{DSE_t\} + \{DSF_t\} \\
 &= \left\{ \left[\underbrace{(OP_t + Special_t + INT_t + ONO_t)}_{Pretax_t} + Tax_t \right] + Extra_t + Minpref_t \right\} \\
 &\quad + \{PY_t + FX_t + REV_t + GW_t + OTH_t\}
 \end{aligned} \tag{1}$$

Clean surplus earnings (CSE_t) is equal to dirty surplus earnings or the reported bottom-line earnings from the “profit and loss account” (DSE_t) plus dirty surplus flows (DSF_t).

First, DSE_t can be decomposed into a series of line items based on the context of profit and loss account: operating profits (OP_t), special items ($Special_t$), net interest expense (INT_t), other nonoperating incomes (ONO_t), tax expense (Tax_t), extraordinary items ($Extra_t$), and minority interests and preference dividends ($Minpref_t$). The introduction of FRS 3: Reporting Financial Performance in 1992 requires three kinds of exceptional items, “gains and losses on sales of fixed assets of continued and discontinued operations,” “gains and losses on sale or termination of continued and discontinued operations,” and “reorganisation,” be separately disclosed as special items ($Special_t$) after operating profit.

Among these line items in the profit and loss account, the sum of OP_t , $Special_t$, INT_t , and ONO_t is called pretax profits ($Pretax_t$), which does not take account of tax expense, extraordinary items, minority interests, and preference dividends. $EARN_t$ is another subtotal earnings, the after-tax profits but before the recognition of extraordinary items, minority interest and preference dividends. As FRS 3 redefines extraordinary items and effectively eliminates the use of extraordinary items, the value for $Extra_t$ is expected to be very small.

In addition, DSF_t in Eq. (1) consists of five categories: prior year adjustments (PY_t), foreign exchange translations (FX_t), asset revaluations (REV_t), net goodwill write-offs (GW_t), and other recognized gains/losses (OTH_t). GW_t is recorded in the

“reconciliation of movements in shareholders’ funds” and the other four classes are in the “statement of total recognised gains and losses.”

B. Decomposition Across Accrual and Cash Flow Components

For the second decomposition of clean surplus earnings across accrual and cash flow components, each line item earnings needs to be decomposed into its corresponding accrual and cash flow component.

Start from the first line item, operating profits:

$$\begin{aligned} OP_t &= [ROCF_t] + [OA_t] \\ &= [ROCF_t] + [WCA_t + LOA_t] \\ &= [ROCF_t] + [(\Delta AR_t + \Delta INV_t - \Delta AP_t) + (DEP_t + OOA_t)] \end{aligned}$$

Operating profits can first be decomposed into reported operating cash flows ($ROCF_t$) and operating accruals (OA_t) since FRS1 (cash flow statement) requires $ROCF_t$ to be reconciled between OP_t and OA_t . Then, I decompose OA_t into working capital accruals (WCA_t) and long-term operating accruals (LOA_t) based on the liquidity. WCA_t has three components: change in accounts receivable (ΔAR_t), change in inventory (ΔINV_t), and change in accounts payable (ΔAP_t). LOA_t consists of depreciation (DEP_t) and other operating accruals (OOA_t).

For the rest of line items, I assume that special items ($Special_t$) and extraordinary earnings ($Extra_t$) are pure accrual components. Without loss of insight and with the added convenience of illustration, I assume that net interest expense (INT_t), tax (Tax_t), and “minority interest plus preference dividends” are cash flow components.¹⁰ I also assume that other nonoperating incomes (ONO_t) and dirty surplus flows (DSF_t) are purely accrual components. Based on these assumptions, I can derive two aggregate figures for total cash flows (TCF_t) and total accruals ($TACC_t$), the sum of which equals clean surplus earnings.

$$\begin{aligned} CSE_t &= [TCF_t] + [TACC_t] \\ &= [ROCF_t + INT_t + TAX_t + Minpref_t] + \{[WCA_t + LOA_t] + Special_t + ONO_t + Extra_t + DSF_t\} \\ &= [ROCF_t + INT_t + TAX_t + Minpref_t] \\ &\quad + \{[(\Delta AR_t + \Delta INV_t - \Delta AP_t) + (DEP_t + OOA_t)] + Special_t + ONO_t + Extra_t + DSF_t\} \end{aligned} \quad (2)$$

Specifically, total cash flows equal reported operating cash flows ($ROCF_t$) plus net interest (INT_t) minus tax expense (TAX_t); total accrual equals the sum of working

¹⁰ Apart from operating cash flow, FRS1 requires the separate disclosure of cash flow for net interests and tax payments. The possibility that this assumption can double count the two items as cash flow components is minimal, given that the tax and net interest are not included in the operating cash flows.

capital accruals (WCA_t : change in accounts receivables plus change in inventory minus change in accounts payable), long-term operating accruals (LOA_t : depreciation and other operating accruals), special items ($Special_t$), extraordinary items ($Extra_t$), and dirty surplus flows (DSF_t). Clean surplus earnings can thus equal all accrual and cash flow components.

79.3.2 The Asymmetric Timeliness Model for Earnings Components

Following Basu (1997), I use a contemporaneous reverse regression¹¹ to assess the asymmetry and timeliness in incorporating good and bad news in earnings. This reverse regression specifies earnings as the dependent variable and the news measure as the independent variable, and then incorporates asymmetry in accounting income timeliness for negative economic news.

$$X_{i,t} = \alpha_1 + \alpha_2 DUM_{MR,t} + \beta_1 MR_t + \beta_2 MR_t \times DUM_{MR,t} + \varepsilon$$

where $X_{i,t}$ is earnings component i at time t ; MR_t is the cum-dividend dollar returns in year t , the surrogate for economic news. $DUM_{MR,t}$ is a dummy variable that takes the value one where MR_t is negative, and zero otherwise. All variables are scaled by opening market value.

The coefficient β_1 captures the responsiveness of the earnings construct ($X_{i,t}$) to the good news, and the coefficient β_2 captures the incremental responsiveness of the earnings construct to contemporaneous bad news. As conservatism can induce asymmetry in earnings timeliness, β_2 is expected to be significantly positive.

Notice that the “timeliness” coefficient is determined by the covariance of accounting earnings and stock returns, divided by the variance of stock returns.

$$\beta_1^{CSE} = \frac{COV(CSE_t, MR_t)}{\sigma_{MR_t}^2} \quad \text{when } MR_t > 0$$

$$\beta_1^{CSE} + \beta_2^{CSE} = \frac{COV(CSE_t, MR_t)}{\sigma_{MR_t}^2} \quad \text{when } MR_t < 0$$

¹¹ One of the reasons to use a reverse regression is to facilitate the exploration of the extent to which accounting earnings can capture the contemporaneously economic performance. Besides, some econometric benefits arise from this reverse regression. As the measurement error of earnings can bias downwards the coefficient of ERC and the R^2 in the normal price-earnings regression, in the reverse regression, the measurement errors will be placed in the disturbance term rather than in the explanatory term, and thus the coefficient will be measured with less errors.

If the dependent variable (aggregated earnings) is decomposed into a series of earnings components, given that the denominator (the variance of stock returns) is constant, the “timeliness” coefficients in each of the subsumed components are additive. The asymmetric timeliness of earnings in clean surplus earnings can be exactly identical to the sum of asymmetric timeliness of earnings in all subsumed components.

Specifically, for the decomposition of clean surplus earnings across line items, the identical relationship between asymmetric timeliness in clean surplus earnings and the sum of asymmetric timeliness of earnings in all line items can be expressed as:

$$\begin{aligned}\beta_1^{CSE} &= \frac{COV[CSE_t, MR_t]}{\sigma_{MR_t}^2} \\ &= \frac{COV[(DSE_t + DSF_t), MR_t]}{\sigma_{MR_t}^2} \\ &= \frac{COV[(OP_t + Special_t + INT_t + ONO_t + Tax_t + Extra_t + Minpref_t) + [PY_t + FX_t + REV_t + GW_t + OTH_t]), MR_t]}{\sigma_{MR_t}^2} \\ &= [\beta_{1,t}^{OP} + \beta_{1,t}^{Special} + \beta_{1,t}^{INT} + \beta_{1,t}^{ONO} + \beta_{1,t}^{Tax} + \beta_{1,t}^{Extra} + \beta_{1,t}^{Minpref}] + [\beta_{1,t}^{PY} + \beta_{1,t}^{FX} + \beta_{1,t}^{REV} + \beta_{1,t}^{GW} + \beta_{1,t}^{OTH}]\end{aligned}$$

Likewise, β_2^{CSE} can be expanded in the same manner.

$$\begin{aligned}\beta_2^{CSE} &= [\beta_{2,t}^{OP} + \beta_{2,t}^{Special} + \beta_{2,t}^{INT} + \beta_{2,t}^{ONO} + \beta_{2,t}^{Tax} + \beta_{2,t}^{Extra} + \beta_{2,t}^{Minpref}] \\ &\quad + [\beta_{2,t}^{PY} + \beta_{2,t}^{FX} + \beta_{2,t}^{REV} + \beta_{2,t}^{GW} + \beta_{2,t}^{OTH}]\end{aligned}$$

On the other hand, for the decomposition of clean surplus earnings across cash flow and accrual components, the identical relationship between asymmetric timeliness in clean surplus earnings and the sum of asymmetric timeliness of earnings in all line items can be expressed as:

$$\begin{aligned}\beta_1^{CSE} &= [\beta_{1,t}^{ROCF} + \beta_{1,t}^{INT} + \beta_{1,t}^{TAX} + \beta_{1,t}^{Minprf}] \\ &\quad + \left[(\beta_{1,t}^{\Delta AR_t} + \beta_{1,t}^{\Delta INV_t} - \beta_{1,t}^{\Delta AP_t}) + (\beta_{1,t}^{DEP_t} + \beta_{1,t}^{OOA_t}) + \beta_{1,t}^{Special} + \beta_{1,t}^{ONO} + \beta_{1,t}^{Extra} + \beta_{1,t}^{DSF} \right] \\ \beta_2^{CSE} &= [\beta_{2,t}^{ROCF} + \beta_{2,t}^{INT} + \beta_{2,t}^{TAX} + \beta_{2,t}^{Minprf}] \\ &\quad + \left[(\beta_{2,t}^{\Delta AR_t} + \beta_{2,t}^{\Delta INV_t} - \beta_{2,t}^{\Delta AP_t}) + (\beta_{2,t}^{DEP_t} + \beta_{2,t}^{OOA_t}) + \beta_{2,t}^{Special} + \beta_{2,t}^{ONO} + \beta_{2,t}^{Extra} + \beta_{2,t}^{DSF} \right]\end{aligned}$$

To summarize, the asymmetric timeliness in clean surplus earnings can be appropriately disaggregated into a series of asymmetric timeliness across line items and across cash flow and accrual components. This innovation is helpful in investigating the contribution made by each earnings component in the asymmetric timeliness and therefore can determine the sources of asymmetric timeliness of earnings.

79.4 Sample Selection

The data used for the tests consists of all industrial¹² firms based in the UK during 1992–2000. All data in the study is collected mainly from DATASTREAM and supplemented with published financial statements. The reason to set the starting year as 1992 is that the provision of complete components of changes in shareholders' funds in DATASTREAM starts since the introduction of FRS3¹³ in 1992.

Besides, samples extracted from DATASTREAM include both live and dead firms to avoid the survivorship bias. An important feature of my data is that all reported incomes, dirty surplus flows, net dividends, and book values of shareholders' funds should be connected up in a transparent way. This ensures that all accounting flows are considered in the analysis. Thus, I firstly collect the complete components of "Reconciliation of Movements in Shareholders' Funds" from DATASTREAM, including net incomes, dirty surplus flows, and opening (ending) book value of equity. As some discrepancies or gaps on changes in shareholders' funds are observed in some data collected from DATASTREAM, I need to conduct a validation investigation in such discrepancies,¹⁴ by reference to published financial statements, to ensure a fully articulated data on net income, dirty surplus flows, net dividends, and book values. In most cases, the discrepancies can be resolved and corrected satisfactorily. If it is not likely to resolve the discrepancies, I will delete the corresponding firms from the samples. This process results in the loss of 64 firms and 353 firm/year observations, approximately 4% of the initial samples.

Moreover, the study also requires additional data on accrual components, cash flow components, and line items. Exhibit 2 summarize the definition of the earnings components and the corresponding Datastream (or Worldscope) codes drawn from 1992 to 2000.

In order to conduct comparative analysis among earnings components, the sample includes only the observations when all the earnings components are available in DATASTREAM for a particular year, and the sum of all line items equals reported earnings.¹⁵ To mitigate the potential effects of outliers, I also delete observations in the top and bottom 1% of each year's distribution of each variable. Such screening

¹²I exclude financial firms because I expect them to exhibit a significant different association between earnings and market returns than do nonfinancial firms. For example, financial firms usually apply fair value accounting to a greater extent than nonfinancial firms and could result in a higher intensity in applying asset impairment rules.

¹³*Financial Reporting Standards 3: reporting financial performance*. FRS3 requires that the information about change in shareholders' funds should be disclosed in a supplementary statement, "Reconciliation of Movements in Shareholders' Funds."

¹⁴For example, the discrepancy between "ending book value of shareholders' funds" in the previous period and the "opening book value of shareholders' funds" in current period; the discrepancy between "the sum of opening book value of shareholders' funds and current-period movements in shareholders' funds" and "the ending book value of shareholders' funds."

¹⁵The non-missing variables requirements might cause some survivorship bias, but some analysis shows that my sample is economically representative of the population both in terms of the market value distribution and industry composition.

Exhibit 2 Accounting variable collection

Line items		Accrual components		
Variables	DATASTREAM code	Variables	DATASTREAM/or Worldscope code ^a	
Operating profit (<i>OP_t</i>)	DS#993	Operating accruals (<i>OA_t</i>)	DS#993- DS#1015	
Special items (<i>Special_t</i>)	DS#1083	Working capital ($WCA_t = \Delta AR_t + \Delta INV - \Delta AP_t$):	$\Delta WS\#4825 + \Delta WS\#4826 - \Delta WS\#4827$	
Special items-termination	DS#1079	Accounts receivable (<i>AR_t</i>)	WS#4825	
Special items-sale of property	DS#1081	Inventory (<i>INV_t</i>)	WS#4826	
Special items-others	DS#1082	Accounts payable (<i>AP_t</i>)	WS#4825	
Net interest expense (<i>INT_t</i>)	DS#2408	Long-term operating accruals ($LOA_t = OA_t - WC_t$):	(DS#993-DS#1015)-(WS#4825 + Δ WS#4826 - $\Delta WS\#4827$)	
Other nonoperating incomes (<i>ONO_t</i>)	DS#154-(DS#993 + DS#1083- DS#2408)	Depreciation (<i>DEP_t</i>)	WS#1148	
Pretax profits (<i>Pretax_t</i>)	DS#154	Other operating accruals ($OOA_t = LOA_t - DEP_t$)	(DS#993-DS#1015)-(WS#4825 + Δ WS#4826 - $\Delta WS\#4827$)- WS#1148	
Taxation (<i>Tax_t</i>)	DS#203			

Net profit before extra and minority ($EARNI_t$)	DS#623	Cash flow components	
Extraordinary items ($Extra_t$)	DS#193	Variables	DATASTREAM code
Minority interest and preference dividends	DS#176 + DS#181	Reported operating cash flows ($ROCF_t$)	DS#1015
Dirty surplus earnings (DSE_t)	DS#1087		
Dirty surplus flows (DSF_t) [=prior year adjustment (PYA_t) + foreign exchange translation (FX_t) + revaluation (REV_t) + goodwill write-offs (GW_t) + other flows (OTH_t)]	$DSF = PYA + DS\#1098 + DS\#1099 + (DS\#1102 - DS\#1103) + (DS\#1100)$		
Clean surplus earnings (CSE_t)	$DS\#1087 + DSF$		

^aThe replacement of DataStream Company Accounts data with the Worldscope standardized Company Accounts set occurred in April 2004. Since then, the Definitions of the Worldscope items have replaced DataStream accounts data items. If items collected after April 2004, it would require the use of Worldscope definitions

Table 1 Sample collection and variables

	Firms	Firm-year observations
Number of firms and firm-year observations		
The initial collection from DATASTREAM (1992–2000)	1704	8917
Less: firms that are difficulty to investigate the discrepancies in the items, and that have unusual items.	(64)	(353)
Less: firms employ merger accounting	(20)	(128)
Initial observations for the estimation	1620	8436
Less: observations with incomplete line items and observations in the top and bottom 1% of each year's distribution of each line items	(117)	(1715)
Total observations for the estimation	1503	6721

processes reduce the firm-year observations from 8430 to 6721 in this chapter. Table 1 displays the details for all the data collected.

79.5 Empirical Results

79.5.1 Summary Statistics

Table 2 contains the descriptive statistics and correlation matrix for the main variables pooled across time. I delete observations in the top and bottom 1% of each year's distribution of each variable to reduce the effects of outliers. I also deflate all variables by the market value of equity in the beginning of the fiscal year in order to reduce heteroscedasticity in the data. The table provides mean, quartiles, and standard deviation for all deflated earnings constructs, across 1992–2000 and also provides the statistics for news measures.

First, as the earnings variables are displayed in the context of the profit and loss accounts, the mean value can add up across all earnings components. For example, the mean value of pretax earnings (Pretax) equals the sum of the mean values of operating profits (OP), special items (Special), net interest expense (INT), and other nonoperating incomes (ONO); the mean value of clean surplus earnings (CSE) equals the sum of the mean value of Pretax, tax, extraordinary items (Extra), minority interest and preference dividends (Minpref), and dirty surplus flows (DSF). In addition, consistent with previous research in conservatism (Basu 1997; Ball et al. 2000), most of the earnings constructs are negatively skewed (all medians exceed means), reflecting a large number of firms reporting large losses. The mean value of OP, Pretax, Earn, DSE, and CSE (0.09, 0.07, 0.043, 0.040, and 0.02) is smaller than the corresponding median value (0.10, 0.10, 0.07, 0.07, and 0.05). This suggests that large one-time negative transitory items lower the mean of earnings variables.

Another point to note is that the extraordinary items are effectively eliminated since FRS3, as 99.9% of firm-year observations are zero.

Table 2 Descriptive statistics and correlations on the variables during 1992–2000

Panel A: Descriptive statistics										
	N	Mean	St. Dev.	Min	Q1	Median	Q3	Max	% = 0	% > 0
Earnings										
<i>OP</i>	6721	0.098	0.146	−0.898	0.057	0.109	0.162	1.035	0	85.3
<i>Special</i>	6721	−0.006	0.093	−1.452	0	0	0	1.618	57.4	23.2
<i>INT</i>	6721	−0.020	0.049	−1.308	−0.028	−0.009	0.0004	0.129	0.4	26.3
<i>ONO</i>	6721	0.001	0.184	−0.477	0	0	0	0.446	62.0	23.7
<i>Pretax</i>	6721	0.073	0.177	−1.397	0.042	0.101	0.148	1.396	0	82.6
<i>Tax</i>	6721	−0.030	0.031	−0.412	−0.043	−0.029	−0.010	0.359	8.2	14.4
<i>EARN</i>	6721	0.043	0.160	−1.351	0.027	0.071	0.106	1.210	0	81.5
<i>Extra</i>	6721	0.00003	0.006	−0.327	0	0	0	0.298	99.9	0.1
<i>Minpref</i>	6721	−0.003	0.017	−0.756	−0.0008	0	0	0.209	57.8	63.1
<i>DSE</i>	6721	0.040	0.159	−1.351	0.025	0.069	0.103	1.210	0	81.0
<i>DSF</i>	6721	−0.019	0.159	−1.919	−0.017	0	0.001	1.694	22.1	30.2
<i>CSE</i>	6721	0.021	0.201	−1.678	−0.010	0.054	0.098	1.649	0	72.9
<i>DSF</i>										
<i>PYA</i>	6721	−0.0002	0.047	−1.220	0	0	0	1.612	90.7	3.4
<i>FX</i>	6721	−0.002	0.027	−0.443	−0.001	0	0	0.909	44.1	22.7
<i>REV</i>	6721	0.003	0.040	−0.541	0	0	0.836	0.102	85.9	8.3
<i>GW</i>	6721	−0.020	0.144	−2.087	−0.005	0	0	1.674	54.2	14.7
<i>OTH</i>	6721	0.0002	0.016	−0.166	−0.000001	0	0	0.805	31.5	35.2

(continued)

Table 2 (continued)

Panel A: Descriptive statistics										
	N	Mean	St. Dev.	Min	Q1	Median	Q3.	Max	% = 0	% > 0
Cash flows										
ROCF	6721	0.094	0.157	-0.870	0.031	0.091	0.154	0.919	0	82.4
Accruals										
WCA	6721	-0.017	0.393	-10.61	-0.076	-0.006	0.055	5.075	0	46.3
Dep	6721	0.058	0.066	0	0.021	0.040	0.072	1.132	0.2	99.8
OOA	6721	-0.038	0.440	-5.028	-0.129	-0.026	0.052	9.862	0	40.5
TACC	6721	-0.020	0.215	-1.870	-0.072	-0.006	0.045	1.632	0	46.6
News										
MR	6721	0.135	0.582	-3.028	-0.193	0.045	0.345	4.288		
MR > 0	3695									
MR < 0	3026									

Where *OP* denotes operating profits; *Special* denotes special items; *INT* denotes net interest expense; *ONO* denotes other nonoperating incomes; *Pretax* denotes pretax earnings; *Tax* denotes tax expense; *EARN* denotes net profits before extraordinary items, minority interest and preference dividends; *Extra* denotes extraordinary items; *Minpref* denotes minority interest and preference dividends; *DSE* is reported earnings; *PYA* denotes prior year adjustments; *FX* denotes foreign exchange translations; *REY* denotes asset revaluations; *GW* denotes goodwill write-offs; *OTH* denotes other recognized gains/losses for the year; *DSF* is dirty surplus flows; *CSE* is the clean surplus earnings; *ROCF* is reported operating cash flow; *WCA* is current operating accruals; *OOA* is other operating accruals; *DEP* is depreciation; *TACC* is total accruals, inclusive of dirty surplus flows

Panel B: Pearson (spearman) correlations coefficients are above (below) the diagonal, and two tailed p values are in parentheses

	<i>OP</i>	<i>Special</i>	<i>NOP</i>	<i>Tax</i>	<i>DSE</i>	<i>DSF</i>	<i>CSE</i>	<i>ROCF</i>	<i>WCA</i>	<i>OOA</i>	<i>DEP</i>	<i>TACC</i>	<i>MR</i>
<i>OP</i>		0.036 (<0.01)	-0.192 (<0.01)	0.602 (<0.01)	0.733 (<0.01)	-0.076 (<0.01)	0.521 (<0.01)	0.537 (<0.01)	-0.115 (<0.01)	0.220 (<0.01)	0.155 (<0.01)	0.244 (<0.01)	0.214 (<0.01)
<i>Special</i>	0.055 (<0.01)		0.061 (<0.01)	0.187 (<0.01)	0.596 (<0.01)	-0.226 (<0.01)	0.294 (<0.01)	0.017 (0.15)	-0.068 (<0.01)	0.069 (<0.01)	-0.048 (<0.01)	0.276 (<0.01)	0.075 (<0.01)
<i>NOP</i>	-0.301 (<0.01)	0.049 (<0.01)		0.040 (<0.01)	0.205 (<0.01)	-0.066 (<0.01)	0.111 (<0.01)	-0.080 (<0.01)	0.041 (<0.01)	-0.004 (<0.01)	-0.442 (<0.01)	-0.061 (<0.01)	0.043 (<0.01)
<i>Tax</i>	0.735 (<0.01)	0.109 (<0.01)	0.033 (<0.01)		0.461 (<0.01)	-0.079 (<0.01)	0.303 (<0.01)	0.377 (<0.01)	-0.129 (<0.01)	0.164 (<0.01)	0.098 (<0.01)	0.162 (<0.01)	0.254 (<0.01)
<i>DSE</i>	0.825 (<0.01)	0.322 (<0.01)	-0.029 (<0.01)	0.685 (<0.01)		-0.207 (<0.01)	0.629 (<0.01)	0.383 (<0.01)	-0.098 (<0.01)	0.202 (<0.01)	-0.091 (<0.01)	0.331 (<0.01)	0.206 (<0.01)
<i>DSF</i>	-0.102 (<0.01)	-0.057 (<0.01)	-0.017 (0.16)	-0.118 (<0.01)	-0.128 (<0.01)		0.630 (<0.01)	-0.006 (<0.01)	-0.0003 (0.98)	-0.022 (0.11)	0.026 (0.04)	0.599 (<0.01)	-0.046 (<0.01)
<i>CSE</i>	0.608 (<0.01)	0.243 (<0.01)	-0.002 (0.88)	0.503 (<0.01)	0.724 (<0.01)	0.394 (<0.01)		0.299 (<0.01)	-0.078 (<0.01)	0.142 (<0.01)	-0.051 (<0.01)	0.739 (<0.01)	0.128 (<0.01)
<i>ROCF</i>	0.586 (0.638)	0.045 (0.458)	-0.163 (<0.01)	0.4700 (<0.01)	0.516 (<0.01)	-0.048 (<0.01)	0.399 (<0.01)		-0.0004 (0.98)	-0.245 (<0.01)	0.405 (<0.01)	-0.359 (<0.01)	0.118 (<0.01)
<i>WCA</i>	-0.104	-0.008	-0.007	-0.079	-0.088	0.023	-0.082	-0.010		-0.919	-0.017	-0.103	-0.080

(continued)

Panel B: Pearson (spearman) correlations coefficients are above (below) the diagonal, and two tailed p values are in parentheses

	<i>OP</i>	<i>Special</i>	<i>NOP</i>	<i>Tax</i>	<i>DSE</i>	<i>DSF</i>	<i>CSE</i>	<i>ROCF</i>	<i>WCA</i>	<i>OOA</i>	<i>DEP</i>	<i>TACC</i>	<i>MR</i>
	(<0.01)	(0.56)	(0.61)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(0.45)		(<0.01)	(0.20)	(<0.01)	(<0.01)
<i>OOA</i>	0.146	0.021	0.069	0.132	0.135	-0.049	0.119	-0.292	-0.773		-0.233	0.323	0.092
	(<0.01)	(0.13)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)		(<0.01)	(<0.01)	(<0.01)
<i>DEP</i>	0.335	-0.049	-0.436	0.177	0.174	-0.007	0.097	0.488	-0.030	-0.302		-0.193	0.031
	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(0.02)	(<0.01)		(<0.01)	(0.02)
<i>TACC</i>	0.189	0.204	-0.0008	0.180	0.252	0.365	0.547	-0.353	-0.075	0.394	-0.237		0.058
	(<0.01)	(<0.01)	(0.94)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)		(<0.01)
<i>MR</i>	0.356	0.086	0.080	0.348	0.382	-0.061	0.266	0.237	-0.125	0.136	0.021	0.065	
	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(0.11)	(<0.01)	

Where *OP* denotes operating profits; *Special* denotes special items; *NOP* denotes nonoperating incomes (interest expense plus other operating incomes); *Tax* denotes tax expense; *DSE* is dirty surplus flows; *DSF* is dirty surplus flows; *CSE* is the clean surplus earnings; *ROCF* is reported operating cash flow; *WCA* is working capital accruals; *OOA* is other operating accruals; *DEP* is depreciation; *TACC* is total accruals, inclusive of dirty surplus flows; *MR* is money returns

Moreover, for the operating cash flow variable, the difference between the mean (0.094) and median (0.091) in the operating cash flows is not as evident as other earnings components. However, the values of accruals components (i.e., special, WC, OOA, TACC) are not only left-skewed but also negative. For example, the mean value of working capital (WC), other operating accruals (OOA), and total accruals (TACC) are -0.017 , -0.038 , and -0.020 , respectively, on average negative, and are much lower than the corresponding median values, -0.006 , -0.026 , and -0.006 , respectively. This supports the conservative accounting theory that the timely recognition of bad news in earnings is driven by the accrual component rather than the cash flows component of earnings since the negative skewness in the distribution of earnings components indicates accounting conservatism.

Next, inspection of the mean value of total dirty surplus flows (DSF: -0.019) and inspection of the mean value in each dirty surplus flow show that most of the dirty surplus flows tend to be negative and are generally quite small. Nonzero dirty surplus flows are relatively infrequent: PYA occurs in approximately 10% of firm-year observations, FX in 55%, REV in 14%, and GW in 45%. The mean value of goodwill flows is -0.020 . The magnitude relative to other dirty surplus also suggests that it was the main component of UK dirty surplus flows before FRS10 eliminated dirty surplus treatment of goodwill write-off in 1998. As the mean value of revaluation is positive, the upward revaluation to tangible assets is performed across UK firms.

Finally, Panel A also shows that the return measure is positively skewed. The left skewness in earnings constructs and the right skewness in returns suggest the existence of asymmetry for news recognition in earnings components.

Panel B presents the Pearson and Spearman correlation between the key variables. Note that the correlation between returns and all variables except dirty surplus flows and current operating accruals are greater than zero. Operating profits are significantly positively correlated with operating cash flows and negatively correlated with current operating accruals. Also, operating cash flow is negatively correlated with current accruals, other operating accruals, and total accruals.

79.5.2 Tests for H1a: Asymmetric Timeliness in Operating Cash Flow and Accrual Components

To explore the relation between asymmetric timeliness of earnings and earnings conservatism, this section disaggregates earnings into cash flow components and accrual components, and employs those components in the Basu model. The first hypothesis predicts that the asymmetric timeliness can only be observed in accrual components, but not in operating cash flow.

Table 3 presents the results for the pooled cross-sectional regression in the earnings components of clean surplus earnings, with the dummy variables capturing the intercept and slope effects for the bad news sample. The sign of raw returns is used as an indicator of good and bad news.

Net profit before extra and minority (<i>EARN_t</i>)	0.074	0.001	-0.002	0.250	10.79	=	0.071	-0.010	-0.020	0.158	4.22	0.003	0.011	0.018	0.092	1.86
	(22.02) **	(0.14)	(-0.2)	(20.29) **			(22.70) **	(-2.00) *	(-4.21) **	(13.17) **		(0.77)	(1.87)	(3.45) **	(6.52) **	
Extraordinary items (<i>Extra_t</i>)	0.0004	-0.000	-0.001	-0.001	-0.04	=						0.0004	-0.000	-0.001	-0.001	-0.04
	(0.59)	(-0.31)	(-0.12)	(-0.21)								(0.59)	(-0.31)	(-0.12)	(-0.21)	
Minority and preference dividend (<i>Minpref_t</i>)	-0.003	-0.002	0.002	-0.004	0.02	=	-0.003	-0.002	0.002	-0.004	0.02					
	(-0.10)	(-1.88)	(0.06)	(-1.47)			(-0.10)	(-1.88)	(0.06)	(-1.47)						
Dirty surplus income (<i>DSE_t</i>)	0.071	-0.001	-0.001	0.245	10.77	=	0.068	-0.012	-0.018	0.154	4.18	0.003	0.011	0.017	0.091	1.86
	(21.93) **	(-0.05)	(-0.24)	(20.171) **			(22.59) **	(-1.84)	(-4.62) **	(13.30) **		(0.79)	(1.86)	(3.44) **	(6.52) **	
Dirty surplus Flows (<i>DSF_t</i>)	-0.015	0.011	-0.018	0.045	0.34	=						-0.015	0.011	-0.018	0.045	0.34
	(-4.49) **	(1.89)	(-3.78) **	(3.50) **								(-4.49) **	(1.89)	(-3.78) **	(3.50) **	
Clean surplus income (<i>CSE_t</i>)	0.056	0.010	-0.019	0.290	7.07	=	0.068	-0.012	-0.018	0.154	4.18	-0.012	0.022	-0.001	0.136	1.17
	(13.50) **	(1.55)	(-3.20) *	(18.64) **			(22.59) **	(-1.84)	(-4.62) **	(13.30) **		(-2.72) **	(2.93) **	(-0.21)	(7.91) **	

^a*MR* is the cum-dividend returns in year *t*; *DUM_{MR}* is a dummy variable that takes the value one when *MR* is negative, and zero otherwise. All variables are scaled by opening market value. The data consists of all the UK industrial firms from 1992 to 2000. Observations falling in the top or bottom 1% of clean surplus (dirty surplus) earnings and returns in each year are excluded. All t-statistics are estimated at the 1% confidence level

The vertical order is based on the context of profit and loss accounts, starting progressively from operating profits to dirty surplus earnings (reported earnings), and the context of the statement of total recognized gains and losses that records dirty surplus flows. Clean surplus earnings are the sum of operating profits, special items, net interest, nonoperating income, tax, extraordinary items, minority and preference dividends, and dirty surplus flows. By construction, the intercept and slope coefficients in clean surplus earnings is exactly identical to the sum of the corresponding coefficients in the above line items. For example, β_2 of the clean surplus earnings is 0.290, the sum of β_2 in operating profits (0.216), special items (0.040), net interest (0.017), nonoperating income (0.005), tax (−0.030), extraordinary items (0.00001), minority and preference dividends (−0.003), and dirty surplus flows (0.045).

Likewise, as earnings equal accruals and cash flow, in each row, the intercept and slope coefficient in the “earnings” column should equal the sum of the corresponding coefficients in the “cash flow” column and the “accrual” column. For example, in the bottom-line row, β_2 in clean surplus earnings (0.290) is equal to β_2 in total cash flows (0.154) and total accruals (0.136). This “adding-up” feature applies to all rows and all columns, and can facilitate my analysis of the relative contribution of operating cash flows and accrual components to the observed asymmetric timeliness effects in earnings.

In this chapter, I am mostly interested in the parameter, β_2 , the incremental response of each earnings component to bad news relative to good news. Under conservative reporting, β_2 is expected to be significantly positive over all accrual components but negligible in operating cash flow components.

Inconsistent with my prediction in H1a, the results show that asymmetric timeliness is not only related to accrual components but also predominantly influenced by cash flow components. As FRS1 (cash flow statement) requires reported operating cash flow to be reconciled between operating profits and operating accruals, the inspection in the first row of “operating profits” provides a broad picture of whether operating accruals or operating cash flow constitutes asymmetric timeliness. Counterintuitively, β_2 in operating profits, along with β_2 in corresponding cash flow and accrual component are significantly different from zero. β_2 in operating profits is 0.216, 79% of which comes from operating cash flow (0.170) and 21% from operating accrual (0.046). Since operating cash flows should not be distorted or influenced by GAAP, this result is inconsistent with the conservatism explanation that asymmetric timeliness of earnings is mainly captured by the accruals rather than the cash flow component of earnings. The preponderance of asymmetric timeliness in operating cash flow calls into question the view that asymmetric timeliness is due to earnings conservatism.

Furthermore, as I move from the row of operating profits downwards to the measure of clean surplus earnings, the pattern that the operating cash flow components dominate the accrual components in the observed asymmetric timeliness of earnings is consistently detected. As “net interest cash flows (INTt)” and “tax payments (Taxt)” are separately identified from operating cash flows in the UK cash flow statement, without loss of insight and with the added convenience of

illustration, I assume INTt and Taxt are cash flow components.¹⁶ Also, “preference dividends and minority interests (Minpref)” is taken as a wholly cash flow item and other nonoperating incomes (ONO) along with dirty surplus flows (DSF) are regarded as accrual components. In such case, the asymmetric timeliness in total cash flows (0.154) is derived from the sum of the asymmetric timeliness in operating cash flows (0.170), net interests (0.017), tax (−0.03), and preference dividends (−0.003); the asymmetric timeliness of total accruals (0.136) is equal to the asymmetric timeliness in operating accruals (0.046), special items (0.040), other non-operating incomes (0.005), extraordinary items (0.0001), and dirty surplus flows (0.045). The asymmetric timeliness in the operating cash flow components (0.154) and the accruals (0.136) can add up to the asymmetric timeliness in the clean surplus earnings (0.290). Again, the contribution of accruals to the overall relation between clean surplus earnings and returns (47%) is smaller than the contribution of operating cash flows (53%). The results refute the theory that asymmetric timeliness is an exact result of earnings conservatism.

79.5.3 Tests for H1b: Evaluating the Application of Three Prominent Standards that Incorporate Asymmetric Verification Criteria

This subsection provides evidence on the relative importance of various accrual components. As Basu (1997, p. 16) states that the accruals process with regards to asset write-offs and write-downs is more likely to reflect conservatism than others, asymmetric timeliness is expected to cumulate in the earnings components related to specific standards that have asymmetric verification for gains and losses. My second hypothesis thus explores the consistency between specific standards that have explicit asymmetric requirements and the asymmetric timeliness in corresponding components.

I first examine subsets of operating accruals by decomposing operating accruals (OA) into working capital accruals (WCA) and long-term operating accruals. WCA consists of change in accounts receivable, change in accounts payable, and changes in inventory; LOA consists of depreciation and other operating accruals. Along with special items and dirty surplus flows, Table 4 reports all the accrual components in the clean surplus earnings.

Consistent with the “adding-up” feature in Tables 3 and 4 shows that the intercept and slope coefficients in “total accruals” equal to the sum of corresponding coefficients in “operating accruals,” “special items,” “other non-operating income,” “extraordinary items,” and “dirty surplus flows.” The coefficient in “operating accruals (OA)” is also equal to the corresponding coefficient in working capital accrual (WCA) [or the coefficient in changes in accounts receivable (ΔAR) plus the coefficient in changes in inventory (ΔINV) minus changes in accounts payable

¹⁶The results are qualitatively similar if those two line items are taken exclusively as accrual components.

Table 4 Asymmetric timeliness of accrual components in a reverse regression

Y^a	α_1	α_2	β_1	β_2	Adj. R^2
1 Operating accruals (OA):	0.004 (1.37)	0.006 (1.14)	0.014 (3.41)**	0.045 (3.88) **	0.95
(1) WCA^b :	-0.015 (-1.6)	0.030 (1.98)	-0.057 (-3.9) **	0.074 (2.07)*	0.62
$\Delta AR (+)$	-0.008 (-1.7)	0.012 (1.74)	-0.026 (-4.2) **	0.033 (2.22)*	0.66
$\Delta INV (+)$	-0.004 (-1.4)	0.013 (2.37)*	-0.017 (-3.4) **	0.029 (2.39)*	0.58
$\Delta AP (-)$	0.003 (0.56)	-0.005 (-0.47)	0.014 (1.89)	-0.012 (-0.67)	0.11
(2) LOA :	0.019 (1.65)	-0.024 (-1.76)	0.071 (4.91)**	-0.029 (-1.83)	0.95
DEP	0.059 (36.70) *	-0.001 (-0.44)	0.007 (3.08)**	-0.009 (-1.1)	0.21
OOA	-0.040 (-3.84) **	-0.025 (-1.65)	0.064 (4.33)**	-0.020 (-1.55)	0.77
2 Special items (special):	-0.004 (-2.11) *	0.006 (1.98)*	0.006 (2.15)*	0.042 (5.59) **	0.99
(1) <i>Gains/losses from termination of operation</i>	-0.006 (-3.7) **	0.004 (1.52)	0.006 (2.81)**	0.023 (3.34) **	0.54
(2) <i>Gains/losses from sale of fixed assets</i>	0.001 (2.98)*	0.001 (1.60)	0.0007 (0.83)	0.003 (1.31)	0.01
(3) <i>Gains/losses from reorganization</i>	0.0001 (0.20)	0.001 (0.75)	-0.0007 (-0.6)	0.016 (5.75) **	0.67
3 Other nonoperating income (ONO)	0.003 (6.82) **	-0.001 (-1.3)	-0.002 (-3.6) **	0.005 (3.41) **	0.27
4 Extraordinary items (Extra)	0.0004 (0.59)	-0.000 (-0.31)	-0.001 (-0.12)	-0.001 (-0.21)	-0.04
5 Dirty surplus flows (DSF)	-0.015 (-4.49) **	0.011 (1.89)	-0.018 (-3.78) **	0.045 (3.50) **	0.34
=Total accruals in CSE	-0.012	0.022	-0.001 (-0.21)	0.136	1.17

(continued)

Table 4 (continued)

Y^a	α_1	α_2	β_1	β_2	Adj. R^2
	(-2.72) **	(2.93) **		(7.91) **	

^a Y is the accrual measures; MR is the cum-dividend returns in year t ; DUM_{MR} is a dummy variable that takes the value one when MR is negative, and zero otherwise. All variables are scaled by opening market value. The data consists of all the UK industrial firms from 1992 to 2000. Observations falling in the top or bottom 1% of clean surplus (dirty surplus) earnings and returns in each year are excluded. All t-statistics are estimated at the 1% confidence level

^bWhere WCA is working capital accruals; ΔAR is change in accounts receivables; ΔAP is change in accounts payable; ΔINV is change in inventory; DEP is depreciation; LOA denotes long-term operating accruals; OOA is other operating accruals; DSF is dirty surplus flows; CSE is the clean surplus earnings

(ΔAP)] and the coefficient in long-term operating accruals (LOA) [or the coefficient in depreciation (DEP) plus the coefficient in other operating accruals (OOA)]. Likewise, the coefficient in “special items” is identical to the sum of the corresponding coefficient in three subcomponents: “gains /losses from termination of operating,” “gains /losses from sale of fixed assets,” and “gains /losses from reorganisation.”

The results in Table 4 indicate that, across all different level of accruals, the R^2 in the accruals are extremely low and the corresponding asymmetric timeliness effects are quite small. For example, the asymmetric timeliness in the three items of working capital accruals (WCA) is only 0.035 and insignificant; the effect of depreciation (DEP) is also insignificantly small. Consistent with the results in Sect. 79.5.2, the low overall effect of asymmetric timeliness in accrual components casts doubt on the representation of “asymmetric timeliness” for earnings conservatism.

Second, I narrow down my analyses on three prominent standards that incorporate asymmetric specification for gains and losses.

1. LCM Rule: Change in Inventory

SSAP 9 stipulates that stock value should be written down to net realizable value if market value is lower than the recorded value, but not vice versa. Such write-down is usually operated through the decrease of inventory. Table 4 shows that asymmetric timeliness in the changes in inventory (ΔINV) is only 0.024 and not significant. This is in contrast with the prediction that LCM rule should accelerate the timely recognition of bad news and delay the recognition of good news.

2. Impairment Test and Contingent Assets and Liabilities: Exceptional Items

FRS11 specifies the impairment test for tangible assets, intangible assets, and goodwill that assets should be written down if the recoverable amount is less than the

book value, but not vice versa. FRS11 requires that the write-downs of asset value be recorded in exceptional items.

In the UK, there are two types of exceptional items: operating exceptional items and special items. “Operating exceptional items” are disclosed above operating profits that are related to operating activities. This belongs to the component of “other operating accruals (OOA)” in this chapter. “Special items” are the exceptional items that are irrelevant to normal activities and FRS3 requires the separate disclosures after operating profits. FRS3 further specifies that “gains/losses on sales of fixed assets of continued and discontinued operations,” “gains/losses on sale or termination of continued and discontinued operations,” and “gains/losses from reorganisation” should be separately disclosed as special items after operating profit. Therefore, I expect asymmetric timeliness is strong in exceptional items, including “other operating accruals (OOA)” or “special items.”

Nevertheless, the results are mixed. Although Table 4 shows that asymmetric timeliness is not significant in “other operating accruals (OOA),” special items with regards to one-time transitory items in “special items” exhibit asymmetric timeliness of earnings. The effect is significant in special items, but the value of asymmetric timeliness (0.042) is not very large, constituting only 16% of asymmetric timeliness in dirty surplus earnings (0.251) or 14% of clean surplus earnings (0.294).

3. Contingent Asset and Liabilities: Change in Accounts Receivable and Changes in Accounts Payable

Finally, FRS12 is the standard for provision and contingent liabilities. Although this standard does not involve the exercise of write-off and write-down, this standard also has an asymmetric verification requirement for unrealized gains and losses. It requires the provision recognition for future losses but not a pre-recognition of future benefits. The typical example is bad debt provision and warranty expense. Thus, for the application of bad debt provision or warranty provisions, I expect asymmetric timeliness to be pronounced in the working capital accruals such as change in accounts receivable or change in payable. However, Table 4 shows that the effect in both of the components is insignificant. I thus cannot find any evidence with regards to asymmetric timeliness in the application of FRS12.

To summarize, I cannot find much evidence showing that asymmetric timeliness is a consequence of earnings conservatism, with reference to the three standards that have explicit asymmetric verification requirements. I only find some support for the impairment tests as evidenced in the “special items,” though the effect is not very large. All of these results suggest that some factors other than conservatism might cause asymmetric timeliness of earnings.

79.5.4 Tests for H1c: The Role of Dirty Surplus Flows in Asymmetric Timeliness

Finally, the last hypothesis relates to the role of dirty surplus flows in asymmetric timeliness. Dirty surplus flows are the gains and losses not included in the profits and

loss accounts. While it has long been argued that the low transparency of dirty surplus flows can create the loophole for managers to hide bad news, the introduction of “statement of total recognised gains and losses” and “reconciliation of movements in shareholders’ funds” by FRS3 requires the disclosure of dirty surplus flows, which can mitigate the manager’s opportunistic behavior. The hypothesis therefore predicts that asymmetric timeliness in dirty surplus flow should not be pronounced after FRS3.

One way to measure the role of dirty surplus flows (DSF) in clean surplus earnings (CSE) is by comparing the decrease in the adjusted R^2 . The adjusted R^2 decreases from 10.77% for dirty surplus earnings to 7.07% for CSE. This result suggests that dirty surplus flows might create some noise in the earnings-news relationship.

The effects of DSF can also be seen by examining the individual asymmetric timeliness of earnings in five classes of DSF, which is reported in Table 5. It is surprising that the effects in each class of dirty surplus flows are very slight. The explanatory power across five classes of flows is nearly zero, and the incremental coefficient is overall insignificant except goodwill. Among the asymmetric timeliness effects in PYA, FX, REV, GW, and Other (0.0005, -0.003 , 0.001, 0.040, and -0.0009 , respectively), only GW is significant at the 1% level. Before the implementation of FRS10 in 1998, SSAP requires goodwill to be written off against shareholders’ equity. The observed asymmetric timeliness of earnings in goodwill may reflect the fact that acquisition is usually perceived as bad news

Table 5 Asymmetric timeliness of dirty surplus flows in a reverse regression

$Y_t = \alpha_1 + \alpha_2 DUM_{MR,t} + \beta_1 MR_t + \beta_2 MR_t \times DUM_{MR,t} + \varepsilon^a$					
Y	α_1	α_2	β_1	β_2	$Adj.R^2$
<i>PYA</i>	0.0005	0.0001	-0.002	0.005	-0.01
	(0.45)	(0.08)	(-1.16)	(1.23)	
<i>FX</i>	-0.002	0.0002	0.001	-0.000	-0.04
	(-4.17)**	(0.20)	(0.52)	(-0.15)	
<i>REV</i>	0.003	0.001	0.001	0.001	-0.00
	(3.38)**	(0.77)	(1.33)	(0.35)	
<i>GW</i>	-0.017	0.010	-0.019	0.040	0.46
	(-5.48)**	(1.91)	(-4.28)**	(3.44)**	
<i>OTH</i>	0.0005	-0.0004	0.001	-0.001	0.04
	(1.59)	(-0.86)	(1.39)	(-0.65)	
<i>Total DSF</i>	-0.015	0.011	-0.018	0.045	0.34
	(-4.49)**	(1.89)	(-3.78)**	(3.50)**	

^a Y is the measure for dirty surplus flows; MR is the cum-dividend returns in year t ; DUM_{MR} is a dummy variable that takes the value one when MR is negative, and zero otherwise. *PYA* denotes prior year adjustments; *FX* denotes foreign exchange translations; *REV* denotes asset revaluations; *GW* denotes goodwill write-offs; *OTH* denotes other recognized gains/losses for the year; *DSF* is dirty surplus flows. All variables are scaled by opening market value. The data consists of all the UK industrial firms from 1992 to 2000. Observations falling in the top or bottom 1% of clean surplus (dirty surplus) earnings and returns in each year are excluded. All t-statistics are estimated at the 1% confidence level

for bidder shareholders when the takeover is carried out with stock offer (Bruner 2002).

To sum up, the components of dirty surplus flows are generally quite small in relation to clean surplus earnings. The results indicate that dirty surplus flows might simply add volatility into reported earnings, given that no apparent asymmetric timeliness effects exist in most of dirty surplus flows and the R^2 also reduces in clean surplus earnings as compared to dirty surplus earnings. While it is often argued that dirty surplus accounting in the UK can conceal some value-relevant information, particularly bad news being prioritized, the introduction of FRS3 in 1992, which attempts to achieve full articulation of financial performance earnings by means of the statement of total recognized gains and losses, might constrain managers' incentives to disclose bad news in the reserves if all aspects of earnings information are disclosed with equal prominence.

79.6 Further Analysis of New Measures

To examine whether the results shown in Sect. 79.5 are sensitive to the way news measures are determined, I also use excess value created (EVC) as an additional news proxy. EVC is defined as the excess of money returns over the required returns from shareholders. In other words, EVC incorporates shareholders' cost of capital in determining the performance and is suggested to be a better performance measure than money returns (MR) for shareholders. Table 6 provides the results for the decomposition analysis based on the news proxy, EVC. Cost of equity capital is determined by CAPM: 5% risk premium multiplied by firm-specific beta plus 3-month T-bill rate at time t .

I find that results based on EVC are qualitatively similar to those based on news proxy, MR, as shown in Table 3. I hence conclude that the news benchmark do not substantially influence my results.

79.7 Conclusion

Since Basu (1997), the higher response coefficient of earnings to negative returns than positive returns has been attributed to earnings conservatism, and termed as "asymmetric timeliness of earnings" (Hsu et al. 2011, 2012; Chan and Hsu 2013; Chan et al. 2015; Wang et al. 2017). This measure of asymmetric timeliness has been extensively used in assessing accounting conservatism and its economic consequences. Nevertheless, there is some doubt as to the sources of asymmetric timeliness of earnings. In particular, Dietrich et al. (2007) have argued that Basu's results are completely explained by coefficient bias due to an econometric artefact, as the effect can be observed in the data where, by construction, conservatism does not exist.

Table 6 Asymmetric timeliness of earnings components in clean surplus earnings in a reverse regression

$Y_t = \alpha_1 + \alpha_2 DUM_{EVC,t} + \beta_1 EVC_t + \beta_2 EVC_t \times DUM_{EVC,t} + \varepsilon^a$																
Earnings				Cash flows components						+ Accruals components						
Dependent variables (Y)	α_1	α_2	β_1	β_2	Adj. R^2	=	α_1	α_2	β_1	β_2	Adj. R^2	α_1	α_2	β_1	β_2	Adj. R^2
Operating profits (OP_t)	0.138	-0.026	-0.012	0.135	8.82	=	0.132	-0.022	-0.019	0.127	6.35	0.007	-0.004	0.007	0.008	0.57
	(50.41)*	(-6.41)*	(-7.1)*	(18.08)*			(44.00)*	(-4.87)*	* (-10.9)	(15.59)*		(2.42)*	(-1.0)	(4.53)*	(1.07)	
Special items ($Special_t$)	0.0003	0.004	-0.0014	0.045	1.38	=						0.0003	0.004	-0.0014	0.045	1.38
	(0.17)	(1.42)	(-1.4)	(9.01)*								(0.17)	(1.42)	(-1.4)	(9.01)*	
Net interest expense (INT_t)	-0.017	0.002	0.001	0.020	0.98	=	-0.017	0.002	0.001	0.020	0.98					
	(-18.6)*	(1.60)	(0.17)	(7.6)*			(-18.6)*	(1.60)	(0.17)	(7.6)*						
Nonoperating income (NOI_t)	0.001	0.0002	0.0003	0.001	0.02	=						0.001	0.0002	0.0003	0.001	0.02
	(3.98)*	(0.73)	(0.73)	(1.63)								(3.98)*	(0.73)	(0.73)	(1.63)	
Profit pretax ($Pretax_t$)	0.123	-0.020	-0.012	0.201	10.96	=	0.115	-0.020	-0.018	0.147	7.36	0.008	0.0002	0.006	0.054	1.22
	(37.01)*	(-3.98)*	(-6.5)*	(22.37)*			(37.67)*	(-4.25)*	* (-10.4)	(17.9)*		(2.45)*	(0.04)	(3.08)*	(5.75)*	
Tax (Tax_t)	-0.040	0.009	0.001	-0.023	9.18	=	-0.040	0.009	0.001	-0.023	9.18					
	(-68.01)*	(10.5)*	(3.61)*	(-14.38)*			* (-68.01)	(10.5)*	(3.61)*	* (-14.38)						
Net profit before extra and minority ($EARN_t$)	0.083	-0.011	-0.011	0.178	9.70	=	0.075	-0.011	-0.017	0.124	5.52	0.008	0.0002	0.006	0.054	1.17
	(27.40)*	(-2.3)*	(-6.4)*	(21.77)*			(25.57)*	(-2.45)*	* (-10.3)	(15.79)*		(0.04)	(3.08)*	(5.75)*		
Extraordinary items ($Extra_t$)	0.0004	-0.0001	-0.00001	0.00006	-0.04	=						0.0004	-0.0001	-0.00001	0.00006	-0.04

(continued)

Table 6 (continued)

$Y_t = a_1 + a_2DUM_{EVC,t} + \beta_1EVC_t + \beta_2EVC_t \times DUM_{EVC,t} + \varepsilon^a$																	
Dependent variables (Y)	Earnings				Adj. R^2	=	Cash flows components				Adj. R^2	+	Accruals components				Adj. R^2
	a_1	a_2	β_1	β_2			a_1	a_2	β_1	β_2			a_1	a_2	β_1	β_2	
Minority and preference dividend	(0.77) -0.004	(-0.46) 0.00005	(0.20) -0.0004	0.0003	0.21	=	-0.004	0.00005	-0.0004	0.0003	0.21		(0.77) (-0.46)		(0.20)		
(Minpref)	(-9.48)* 0.079	(0.10) -0.011	(-3.8)* -0.011	(0.37) 0.178			(-9.48)* 0.071	(0.10) -0.011	(-3.8)* -0.017	(0.37) 0.124							
Dirty surplus income (DSE _t)	(23.36)* -0.027	(-2.36)* 0.029	(-6.42)* -0.0002	(21.75)* 0.04	9.66 0.55	=	(24.48)* 0.029	(-2.45)* 0.029	(-10.8)* 0.04	(15.86)* 0.04	5.41 0.55	0.008 -0.027	0.0001 -0.0002	0.006 0.04	0.054 0.04		
Dirty surplus flows (DSF _t)	(-8.72)* 0.052	(5.90)* 0.018	(-0.10) -0.011	(4.75)* 0.218								(-8.72)* -0.019	(5.90)* 0.029	(-0.10) 0.006	(4.75)* 0.094		
Clean surplus income (CSE _t)	(13.33)* 0.052	(3.05)* 0.018	(-5.13)* -0.011	(20.88)* 0.218	7.07	=	(24.48)* 0.071	(-2.45)* -0.011	(-10.8)* -0.017	(15.86)* 0.124	5.41	(-4.57)* -0.019	(4.47)* 0.029	(2.40)* 0.006	(8.22)* 0.094		

^aEVC is the excess of money returns over the cost of equity capital, which is determined by CAPM. 3-month T-bill rate at time t plus firm-specific beta times 5% risk premium t; DUM_{ec} is a dummy variable that takes the value one when EVC is negative, and zero otherwise. All variables are scaled by opening market value. The data consists of all the UK industrial firms from 1992 to 2000. Observations falling in the top or bottom 1% of clean surplus (dirty surplus) earnings and returns in each year are excluded. All t-statistics are estimated at the 1% confidence level

Motivated by the prediction that asymmetric timeliness of earnings due to conservatism ought to be more pronounced in certain earnings components than in others, I measure asymmetric timeliness for components of earnings. The study is the first to employ an innovative decomposition analysis to analyze the source of asymmetric timeliness of earnings across all earnings components. I decompose clean surplus earnings across line items and across accrual and cash flow components, and seek evidence as to whether the incidence of asymmetric timeliness across components is consistent with the theory that such asymmetric timeliness is due to accounting conservatism. The innovative feature of the study is that the asymmetric timeliness in aggregated earnings can be disaggregated into asymmetric timeliness in components. The asymmetric timeliness of clean surplus earnings is exactly identical to the sum of asymmetric timeliness in the subsumed components, the characteristics of which is very helpful in exploring the sources of asymmetric timeliness of earnings.

With articulated dataset, where clean surplus earnings are articulated with net dividends and consecutive changes in book value, the results show that asymmetric timeliness of earnings is not directly related with earnings conservatism. When clean surplus earnings are decomposed into “accrual” components and “cash flow” components, the asymmetric timeliness of earnings in “cash flow” components is substantially higher than the asymmetric timeliness effect in “accrual” components. The explanatory ability of accrual components to reflect contemporaneous good and bad news measures is also extremely low. This is inconsistent with the conservatism proposition that accrual components drive accounting conservatism. The decomposition across line items also finds little evidence with regard to the application of three prominent standards that incorporate asymmetric verification criteria.

These results call into question whether asymmetric timeliness of earnings is a consequence of conservatism. The study provides important implications to accounting researchers who apply the measure in the conservatism study. In particular, the study in the comparison of conservatism across different governance systems and accounting regimes (Ball et al. 2000, 2003) perhaps needs a second look when the interpretation generally given to the governance system or accounting regime with higher asymmetric timeliness of earnings is that the accounting regime is fundamentally more conservative.

Appendix 1 The UK Financial Statements

I extract the financial statements of Boots Plc. as of 1999 to demonstrate the presentation formats of the UK “profit and loss accounts,” the “statement of total recognised gains and losses,” and “reconciliation of movements in shareholders’ funds.”

1. Format of the UK profit and loss accounts

	Notes	Before exceptional items 1999 £m	Exceptional items (note 3) 1999 £m	Total 1999 £m
Turnover				
Continuing operations – excluding acquisitions		4894.7	–	4894.7
– acquisitions		17.7	–	17.7
Turnover from continuing operations		4912.4	–	4912.4
Discontinued operations		132.2	–	132.2
Total turnover	1	5044.6	–	5044.6
Operating profit ⊙				
Continuing operations – excluding acquisitions		563.1	(69.4)	493.7
– acquisitions		(1.7)	(6.9)	(8.6)
Operating profit from continuing operations		561.4	(76.3)	485.1
Discontinued operations		2.7	–	2.7
Group operating profit		564.1	(76.3)	487.8
Share of operating loss of joint venture		(1.7)	–	(1.7)
Total operating profit including joint venture	1, 2	562.4	(76.3)	486.1
Profit on disposal of fixed assets ⊙	3			
Continuing operations		–	4.6	4.6
Discontinued operations		–	0.3	0.3
Loss on disposal of businesses	4			
Discontinued operations		–	(318.9)	(318.9)
Profit on ordinary activities before interest		562.4	(390.3)	172.1
Net interest ⊙	5	(1.8)	–	(1.8)
Profit on ordinary activities before taxation ⊙		560.6	(390.3)	170.3
Tax on profit on ordinary activities ⊙	6	(169.9)	23.6	(146.3)
Profit on ordinary activities after taxation ⊙		390.7	(366.7)	24.0
Equity minority interests ⊙		(0.1)	–	(0.1)
Profit for the financial year attributable to shareholders ⊙	7	390.6	(366.7)	23.9
Dividends	8			(214.5)
(Loss)/profit retained				(190.6)

Note 1: ⊙ denotes the line items covered in this chapter

2. Statement of total recognized gains and losses

It is suggested that the profit and loss account is not sufficient on its own as a report of a company's financial performance. FRS3 introduces the statement of total recognized gains and losses to present details of the other elements making up a company's performance.

Statement of total recognised gains and losses	1999 £m
Profit for the financial year attributable to shareholders	23.9
Deficit on revaluation of properties ☆ ◎	(1.4)
Impairment losses on revalued assets ☆ ◎	(1.7)
Currency translation differences on foreign currency net investments ☆ ◎	3.0
Other gains and losses ☆ ◎	0.4
Total recognised gains and losses for the year	24.2
Prior period adjustment (see note 20) ☆ ◎	12.0
Total gains and losses recognised since last annual report	36.2

3. Reconciliation of movements in shareholders' funds

Reconciliation of Movements in Shareholders' Funds	1999 £m
Total recognised gains and losses for the year	24.2
Dividends	(214.5)
New share capital issued (net of expenses)	8.8
Goodwill relating to acquisitions prior to 1st April 1998 ☆ ◎	(1.4)
Goodwill released on disposal of businesses ☆ ◎	312.2
Scrip dividends	-
Net increase in shareholders' funds	129.3
Opening shareholders' funds	1,650.9
Closing shareholders' funds	1,780.2

Note 1: ◎ denotes the line items covered in this chapter

Note 2: ☆ denotes the five classes of dirty surplus flows. There are five classes of dirty surplus flows: (1) asset revaluation, (2) currency translation, (3) other gains and losses, (4) prior year adjustment, (5) goodwill write-offs and disposition. All dirty surplus flows except goodwill are disclosed in the "statement of total recognised gains and losses." Goodwill acquisition and disposition are recorded in the "reconciliation of movements in shareholders' funds."

Note 3: Financial Reporting Standard 10: *Goodwill and Intangible Assets* requires purchased goodwill and intangible assets to be capitalized and amortized through the profit and loss account over their useful economic lives. FRS10 eliminated this dirty surplus treatment of goodwill write-offs in 1998

References

- Accounting Standards Board. 1971. *Statement of Standard Accounting Practice (SSAP) 2: Disclosure of Accounting Policies*. Accounting Standards Board.
- . 1984. *Statement of Standard Accounting Practice (SSAP) 22: Accounting for Goodwill*. Accounting Standards Board.
- . 1988. *Statement of Standard Accounting Practice (SSAP) 9: Stocks and Long-term Contracts*. Accounting Standards Board.
- . 1991. *Financial Reporting Standard (FRS) 1: Cash Flow Statement*. Accounting Standards Board.
- . 1992. *Financial Reporting Standard (FRS) 3: Reporting Financial Performance*. Accounting Standards Board.
- . 1997. *Financial Reporting Standard (FRS) 10: Goodwill and Intangible Assets*. Accounting Standards Board.
- . 1998a. *Financial Reporting Standard (FRS) 11: Impairment of Fixed Assets and Goodwill*. Accounting Standards Board.
- . 1998b. *Financial Reporting Standard (FRS) 12: Provisions, Contingent Liabilities and Contingent Assets*. Accounting Standards Board.
- . 1999. *Financial Reporting Standard (FRS) 15: Tangible Fixed Assets*. Accounting Standards Board.
- Ball, R., S.P. Kothari, and A. Robin. 2000. The effect of international institutional factors on properties of accounting earnings. *Journal of Accounting and Economics* 29: 1–51.
- Ball, R., A. Robin, and J.S. Wu. 2003. Incentives versus standards: Properties of accounting income in four East Asia countries. *Journal of Accounting and Economics* 36 (1-3): 235–270.
- Basu, S. 1995. *Conservatism and the asymmetric timeliness of earnings*. Ph.D. thesis, University of Rochester, Rochester, New York.
- . 1997. The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics* 24: 3–37.
- Beekes, W., P. Pope, and S. Young. 2004. The link between earnings conservatism and board composition: Evidence from the UK. *Corporate Governance: An International Review* 12: 47–59.
- Black, F. 1993. Choosing accounting rules. *Accounting Horizons* 7 (4): 1–17.
- Bruner, R.F. 2002. Does M&A pay? A survey of evidence for the decision-maker. *Journal of Applied Finance*: 7–27.
- Burke, Q., P. Chen, and G.J. Lobo. 2020. Is corporate social responsibility performance related to conditional accounting conservatism? *Accounting Horizons* 34 (2): 19–40.
- Chan, A.L., and A. Hsu. 2013. Corporate pyramids, conservatism and cost of debt: Evidence from Taiwan. *International Journal of Accounting* 48 (3): 390–413.
- Chan, A.L., A. Hsu, and E. Lee. 2015. Mandatory adoption of IFRS and timely loss recognition across Europe: The effect of corporate finance incentives. *International Review of Financial Analysis* 38: 70–82.
- Dhaliwal, D., K.R. Subramanyam, and R. Trezevant. 1999. Is comprehensive income superior to net income as a measure for firm performance. *Journal of Accounting and Economics* 26: 43–67.
- Dietrich, J.R., I. Karl, A. Muller, and E.J. Riedl. 2007. Asymmetric timeliness tests of accounting conservatism. *Review of Accounting Studies* 12: 95–124.
- Ehrbar, A. 1998. *EVA: The real key to creating wealth*. New York: Wiley.
- Giner, B., and W. Rees. 2001. On the asymmetric recognition of good and bad news in France, Germany and the UK. *Journal of Business Finance and Accounting* 28: 1285–1381.
- Givoly, D., and C. Hayn. 2000. The changing time-series properties of earnings, cash flows and accruals: Has financial reporting become conservative? *Journal of Accounting and Economics* 29: 287–320.

- Givoly, D., C. Hayn, and A. Natarajan. 2007. Measuring reporting conservatism. *The Accounting Review* 82 (1): 65–106.
- Hayn, C. 1995. The information content of loss. *Journal of Accounting and Economics* 20: 125–153.
- Hsu, A., J. O'Hanlon, and K. Peasnell. 2011. Financial distress and the earnings-sensitivity-difference measure of conservatism. *ABACUS* 47 (3): 284–314.
- . 2012. The Basu measure as an indicator of conditional conservatism: Evidence from UK earnings components. *European Accounting Review* 21 (1): 87–113.
- Hsu, C., K. Novoselov, and R. Wang. 2017. Does accounting conservatism mitigate the shortcomings of CEO overconfidence. *The Accounting Review* 92 (6): 77–101.
- International Accounting Standards Board. 1998. *International Accounting Standards (IAS) 36: Impairment of Assets*. International Accounting Standards Board.
- Isidro, H., J. O'Hanlon, and S. Young. 2004. Dirty surplus accounting flows: International evidence. *Accounting and Business research* 34 (4): 383–410.
- Johnson, L.T., C.L. Reither, and R.J. Swieringa. 1995. Toward reporting comprehensive income. *Accounting Horizon* 9 (4): 128–137.
- Lipe, R. 1986. The information contained in the components of earnings. *Journal of Accounting Research* 24: 37–64.
- O'Hanlon, J., and P. Pope. 1999. The value-relevance of UK dirty surplus accounting flows accounting flows. *British Accounting Review* 32: 459–482.
- Pope, P., and M. Walker. 1999. International differences in the timeliness, conservatism and classification of earnings. *Journal of Accounting Research* 37: 53–87.
- Pope, P., and M. Walker. 2003. *Ex-ante and ex-post accounting conservatism, asset recognition and asymmetric earnings timeliness*. Working Paper.
- Robinson, L.E. 1991. The time has come to report comprehensive income. *Accounting Horizon*: 107–112.
- Wang, T., H. Lin, W. Hsu, H. Chen, and C. Liu. 2017. Accounting conservatism: A review of the research based on data from Asian countries. *NTU Management Review* 27 (4): 1–48.
- Watts, R.L. 2003. Conservatism in accounting Part I: Explanations and implications. *Accounting Horizon* 17 (3).
- Zhong, Y., and W. Li. 2017. Accounting conservatism: A literature review. *Australian Accounting Review* 81 (27).



Fang-Pang Lin

Abstract

Securities trading is one of the few business activities where a few seconds processing delay can cost a company a big fortune. The growing competition in the market exacerbates the situation and pushes further toward instantaneous trading even in a split second. The key lies on the performance of the underlying information system. Following the computing evolution in financial services, it was a centralized process to begin with and gradually decentralized into a distribution of actual application logic across service networks. Financial services have a tradition of doing most of its heavy lifting financial analysis in overnight batch cycles. In securities trading, however, it cannot satisfy the need due to its ad hoc nature and requirement of fast response. New computing paradigms, such as grid and cloud computing, aiming at scalable and virtually standardized distributed computing resources, are well suited to the challenge posed by capital market practices. Both have been gaining popularity to serve as a production environment for finance services in recent years, which further leads to the recent rise of big data and the success of large-scale machine learning, impacting on broad and diverse domains, such as finance services. In this entry, the core computing competence for financial services is examined. A comparison of grid and cloud will be briefly described. How the underlying algorithm for financial analysis can take advantage of grid and cloud environment is presented. One of the most popular practiced algorithms Monte Carlo simulation is used in our cases study for option pricing and risk management. The various distributed computational platforms are carefully chosen to demonstrate the performance issue for financial services, which also extends to applications of big data and machine learning.

F.-P. Lin (✉)

National Research Laboratories, National Center for High Performance Computing,
Hsinchu, Taiwan

e-mail: fplin@nchl.org.tw

Keywords

Financial service · Grid and cloud computing · Big data · Machine learning · Monte Carlo simulation · Option pricing · Risk management

80.1 Introduction

80.1.1 Information Technology for Financial Services

The finance service industry involves a broad range of organizations such as banks, credit card companies, insurance companies, consumer finance companies, stock brokerages, investment funds, and some government-sponsored enterprises. The industry represents a significant share of the global market. Information technology (IT) in the financial service industry is considered as an indispensable tool for its competitiveness in the market. The IT spending in financial service industry also grows (Calen 2010). In such a scale of market size, the impact directly from the use of advanced IT brings on the financial service industry.

The structure of the industry has changed significantly in the last two decades as companies, which are not traditionally viewed as financial service providers, have taken advantage of opportunities created by technology entering the market. New technology-based services keep emerging. These changes are a direct result of the interaction of technology with the industrial environment, such as economic atmosphere, societal pressures, and the legal/regulatory environment in which the financial service industry operates. The effects of IT on the internal operations, the structure, and the types of services offered by the financial service industry have been particularly profound (Phillips et al. 1984; Hauswald and Marquez 2003; Griffiths and Remenyi 2003). IT technology has been and continues to be both a motivator and facilitator of change in the financial service industry, which ultimately leads to competitiveness of the industry. The change is in particular radical after 1991 when World Wide Web was invented by Tim Berners-Lee and his group for information sharing in the community of high energy physics. It was later introduced to the rest of the world, which subsequently changed the face of how people do business today.

Informational considerations have long been recognized to determine not only the degree of competition but also the pricing and profitability of financial services and instruments. Recent technological progress has dramatically affected the production and availability of information, thereby changing the nature of competition in such informationally sensitive markets. Hauswald and Marquez (2003) investigate how advances in information technology (IT) affect competition in the financial services industry, particularly credit, insurance, and securities markets. Two aspects of improvement in IT are focused – better processing and easier dissemination of information. In other words, two dimensions of technology progress that affects competition in financial services can be defined as advances in the ability to process and evaluate information, and in the ease of obtaining information generated by

competitors. While better technology may result in improved information processing, it might also lead to low-cost or even free access to information through, for example, informational spillovers. They show that in the context of credit screening, better access to information decreases interest rates and the returns from screening. On the other hand, an improved ability to process information increases interest rates and bank profits. Hence, predictions regarding financial claims' pricing hinge on the overall effect ascribed to technological progress. Their results conclude that in general financial markets' informational asymmetries drive profitability.

The viewpoint of Hauswald and Marquez is adopted in this work. Assuming competitors in the dynamics of financial market possess similar capacity, the informational asymmetries can be created sometimes only between seconds and now are possible to be achieved through the outperformance of underlying IT platforms.

80.1.2 Competitiveness Through IT Performance

Following the computing evolution in financial services, it was a centralized process to begin with and gradually decentralized into a distribution of practical trading application logic across service networks. Financial services have a tradition of doing most of its heavy lifting financial analysis in overnight batch cycles. However, in securities trading, it cannot satisfy the need due to its ad hoc nature and requirement of fast response.

New computing paradigms, grid and cloud computing, emerged in the last decade. It was incorporated into the core context of a well-referenced Atkins' report of National Science Board of the USA, namely "Revolutionizing Science and Engineering Through Cyberinfrastructure" (Atkins et al. 2003), which lays down a visionary path for future IT platform development of the world. One way to observe this trend is from statistics of Google Trends in the global search interest over time of key phrases in the finance sector including cluster computing, grid computing, and cloud computing, which represent three main stream computing paradigms in high-performance quantitative analysis, and big data and machine learning, which have been emerging as a major emphasis on data-centric computing with emphasis of insight from data (Fig. 1).

Cluster computing is a group of coupled computers that work closely together so that in many respects they can be viewed as though they are a single computer. They are connected with high-speed local area networks, and the purpose is usually to gain more compute cycles with better cost performance and higher availability. The grid computing aims at virtualizing scalable geographically distributed computing and observatory resources to maximize compute cycles and data transaction rates with minimum cost. Cloud computing is more of a recent development owing to the similar technology used in global information service providers, such as Google, Amazon, etc. The cloud is referred to as a subset of Internet if to be explained in a simplest fashion. Within the cloud, the computers also talk with servers instead of communicating with each other similarly to that of Peer-to-peer computing (Milojicic et al. 2002). There are no definitive definitions for the above terminology.

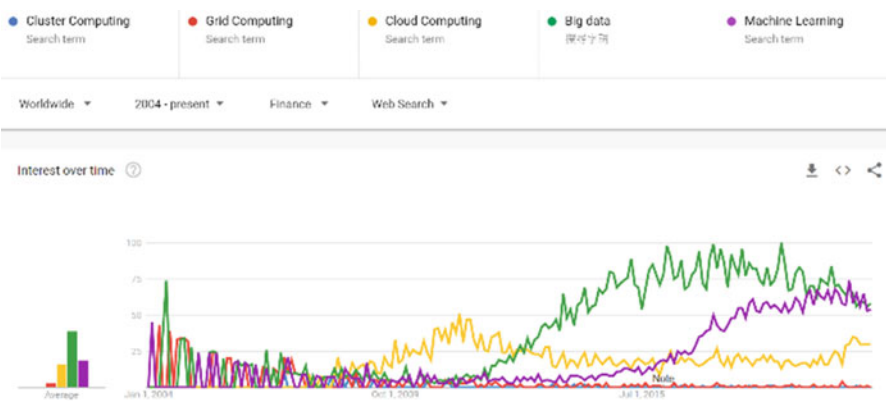


Fig. 1 The trend history on search interest in the finance sector from Google Trends

However, people tend to view clusters as one of foundational components of grids, or grids as a metaccluster on wide area networks. This is also known as horizontal integration. The cloud virtualizes further the compute, store, and network in a utility sense and provides an interface between users and grids. We refer to Foster et al. (2008) and Yelick et al. (2011) for a comparison. This perspective considers grids as a backbone of Cyberinfrastructure to support the clouds. Similarly, in the early days of development of grids, there is a so called “@HOME” style PC grids (Korpela et al. 2001), which are exactly working on at least ten thousands of PCs, in which owners of PCs donate their CPU times when their machines are idle. The PC grids can be specifically categorized as clouds.

Figure 1 shows that cluster and grid computing are of less interest, which also in turn more mature in financial services. Cloud computing since its introduction in mid-2007 still has momentum to develop further. The new rising technology is big data (Bughin et al. 2010), which implies a paradigm shift from compute-centric, network-centric gradually to data-centric computing and has been growing significantly since then. There is also a dramatical increase of interest in machine learning. The phenomenon is still based on the advance of system performance. Clusters are the major market products, either in supercomputers from big vendors, such as Cray, IBM, and NEC, or from an aggregation of computers in university research laboratories. It should be noted that accelerators, such as graphic process unit (GPU), field programmable gate array (FPGA), also play a crucial role in HPC clusters. The introduction of coprocessor plays a key role, which is not discussed here and considered as part of the above compute paradigms. Figure 1 also implies constant market need for high-performance computing. The performance and security issues are fundamental to general distributed and parallel computing, which also remain as a challenge to cluster, grid, cloud, big data, and machine learning (Lauret et al. 2010; Ghoshal et al. 2011; Ramakrishnan et al. 2011). Performance models in compute-based grid and cloud-enabled environment are adopted in this work. The general definitions of grid and cloud computing will be introduced. To tackle the core

performance issue, grids are chosen to demonstrate how fundamental financial calculations can be improved, hence leverage the financial service.

In this chapter, the core computing competence for financial services will be examined and how underlying algorithms for financial analysis can take advantage of grid, big data, and cloud-enabled environment scrutinized. One of the most popular practiced algorithms is Monte Carlo simulation (MCS), and it will be specifically used in our case studies for calculations of option pricing and for value at risk (VaR) in risk management.

Four grid-, Big Data-, and cloud-enabled platforms are carefully chosen to exploit the performance issue for financial services. The first one is traditional grid platform with heterogeneous and distributed resources. Digital packets are usually connected via optical fibers. For long distance, depending on network traffics, it will produce approximately 150 ~ 300 microseconds (mm) latency across Pacific Ocean. This is the physical constrain of light speed when traveling through the fiber channels. Therefore, even in a split second, packets can still travel to anywhere in the world. The Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) grid is a typical example, which linked with 14 countries and 36 sites. The system is highly heterogeneous. The computer nodes mounted to PRAGMA grid range from usual PC clusters to high-end supercomputers. The second one is a special Linux, or DRBL, PC cluster. It converts a system into a homogenous Linux system and exploits the compute cycles of the cluster. The intention is to provide dynamic and flexible resources to cope better with uncertainty of the traders' cycle demand. A PC grid is chosen to demonstrate finance services that can be effectively conducted through a cloud-enabled computing. The usefulness of PC grid is based on the fact that 90% of CPUs' time of PCs was in idle status. Finally, a state-of-the-art Cloud-enabled hybrid high-performance system is introduced, which allows to conduct big data computing as well as the machine learning in scale.

80.2 Performance Enhancement

In this section, two types of grid systems, compute intensive and data intensive, respectively, are introduced. The classification of the types is based on various grid applications. Traditionally, the grid systems provide a general platform to harvest, or to scavenge if used only in idle status, compute cycles for a collection of resources across boundaries of institutional administration. In the real world, most applications are in fact data centric. For example, in a trading center, it collects tick-by-tick volume data from all related financial markets and is driven by informational flows, hence typical data centric. However, as noted in Sect. 80.3, the core competence still lies on the performance enhancement of the IT system. The following two sub-sections give more details of compute-intensive as well as data-intensive grid systems by a survey of the current development of grids specifically for financial services. In some cases, e.g., high frequency data with real-time analysis, two systems have to work together to get better performance. Our emphasis will be more on compute-intensive grid system.

Grid was coined by Ian Foster (Foster and Kesselman 2004). The definition of grid thus far is well accepted and has been stably used up to now. The virtual organization (VO) has strong implication of community-driven and collaborative sharing of distributed resources. The advance of the development of optical fiber network in recent years plays a critical role of why grids can be a reality. It is also the reason why now the computing paradigm has shifted to distributed/grid computing.

Additionally, perhaps the most generally useful definition is that a grid consists of shared heterogeneous computing and data resources networked across administrative boundaries. Given such a definition, a grid can be thought of as both an access method and a platform, with grid middleware being the critical software that enables grid operation and ease-of-use.

The term “cloud computing” has been used to refer to different concepts, models, and services over the last few years. The definition for cloud computing provided by the National Institute of Standards and Technology (NIST) is well received in the IT community, which defines cloud computing as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort of service provider interaction (Mell and Grance 2011). The model gains popularity in the industry for its emphasis on pay-as-you-go and elasticity, the ability to quickly expand and collapse the utilized service as demand requires. Thus, new approaches to distributed computing and data analysis have also emerged in conjunction with the growth of cloud computing. These include models like MapReduce (Dean and Ghemawat 2004) and scalable key-value stores like Big Table (Chang et al. 2006). From the high-performance computing perspective, could computing technology allows users to have the ability to get on-demand access to resources to replace or supplement existing systems, as well as the ability to control the software environment? Yet, the core competence still lies on the performance of financial calculation and further of the transactions of financial processes. This work will focus on the core competence in financial calculation based on grid- and cloud-like environment. Grid computing technology will be used to explain how the core financial calculations can be significantly accelerated in various distributed and parallel computing environments. The calculation models in this work can be easily migrated to pure cloud environments.

80.2.1 Compute-Intensive IT Systems

The recent development of computational finance based on grids is hereby scrutinized and remarks given. Our major interest is to see if the split second performance is well justified under the grid architecture. Also, real-time issue with real market-parametric data should be used as input for practical simulation. In addition, issues of inter-system, inter-disciplinary, geographical distribution of resources and the degree of virtualization are crucial to the success of such a grid. The chosen projects are reviewed and discussed as follows:

80.2.1.1 PicsouGrid

This is a French grid project for financial service. It provides a general framework for computation finance and targets on applications of options trading, options pricing, Monte Carlo simulation, aggregation of statistics, etc. (Stokes-Rees et al. 2007). The key for this development is the implementation of the middleware ProActive. ProActive is an in-house Java library for distributed computing developed by INRIA Sophia Antipolis, France. It provides transparent asynchronous distributed method calls and is implemented on top of Java RMI. It is also used in commercial applications. It also provides fault tolerance mechanism. The architecture is shown in Fig. 2, which is very similar to most of grid applications apart from the software stack used. The option pricing was tested in approximately 894 CPUs. The underlying computer systems are heterogeneous. The system is used for metacomputing. As a result, the system has to specifically design to orchestrate and to synchronize and resynchronize whole distributed processes for one calculation. Once the grid system requires synchronization between processes, which imply stronger coupling of algorithm of interest, the performance will be seriously affected. There is no software treatment to solve such problems and should be tackled by physical infrastructure, e.g., optical fiber network with Layer 2 light path.

80.2.1.2 FinGrid

FinGrid stands for financial information grid. Its study includes components of bootstrapping, sentimental analysis, and multiscale analysis, which focus on information integration and analysis, e.g., data mining. It takes advantage of the huge collection of numerical and textual data simultaneously to emphasize the study of societal issues (Amad et al. 2004; Ahmad et al. 2005; Gillam et al. 2005). The architecture of FinGrid is shown in Fig. 3. It is a typical three-tier system, in which the first tier facilitates the client in sending a request to one of the services: text

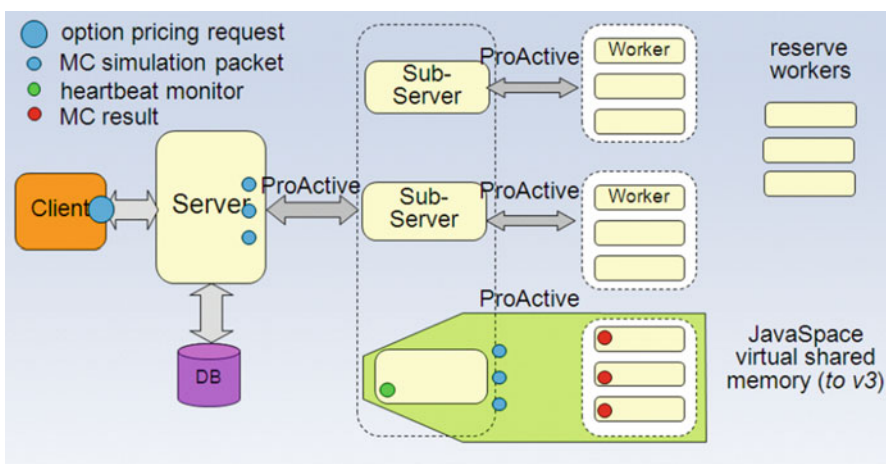


Fig. 2 Architecture of PicsouGrid for option pricing based on Monte Carlo simulation

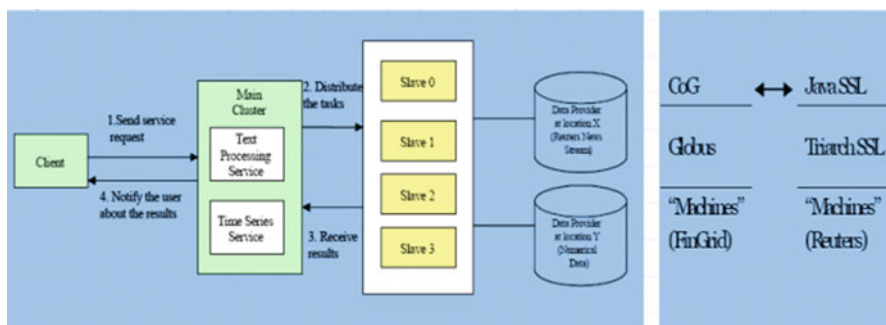


Fig. 3 The architecture of financial information grid

processing service or time series service; the second tier facilitates the execution of parallel tasks in the main cluster and is distributed to a set of slave machines (nodes); and the third tier comprises the connection of the slave machines to the data providers. This work focuses on small-scale and dedicated grid system. It pumps in real and live numerical and textual data from say Reuters and performs real-time sophisticated data mining analysis. This is a good prototype for Financial grid. However, it will encounter similar problems as that of PicasouGrid if it is to scale up. The model is more successful in automatically combining real data and the analysis.

- IBM Japan collaborated with life insurance company and adopted the PC grids concept to scavenge more compute cycles (Tanaka 2003):

In this chapter, an integrated risk management system (see Fig. 4) is modified, in which the future scenarios of the red circle of Fig. 4 are sent via grid middleware to a cluster of PCs. According to the size of the given PCs, the number scenarios are then divided in a work-balanced manner for each PC. This is the most typical use of compute-intensive grid systems and a good practice for production system. However, the key issues that were discussed in the above two cases cannot be answered in this chapter. Similar architecture can also be found in EGrid (Leto et al. 2005).

- UK e-Science developed a grid service discovery in the financial market sector focusing on integration of different knowledge flows (Bell and Ludwig 2005).

From application's viewpoint, business and technical architecture of financial service applications may be segmented by product, process, or geographic concerns. Segmented inventories make inter silo reuse difficult. The service integration model is adopted and a loosely coupled inventory – containing differing explicit capability knowledge. Three use-cases were specifically chosen in this work to explore the use of semantic searching:

Use-case 1 – Searching for trades executed with a particular counterparty.

Use-case 2 – Valuing a portfolio of interest rate-derivative products.

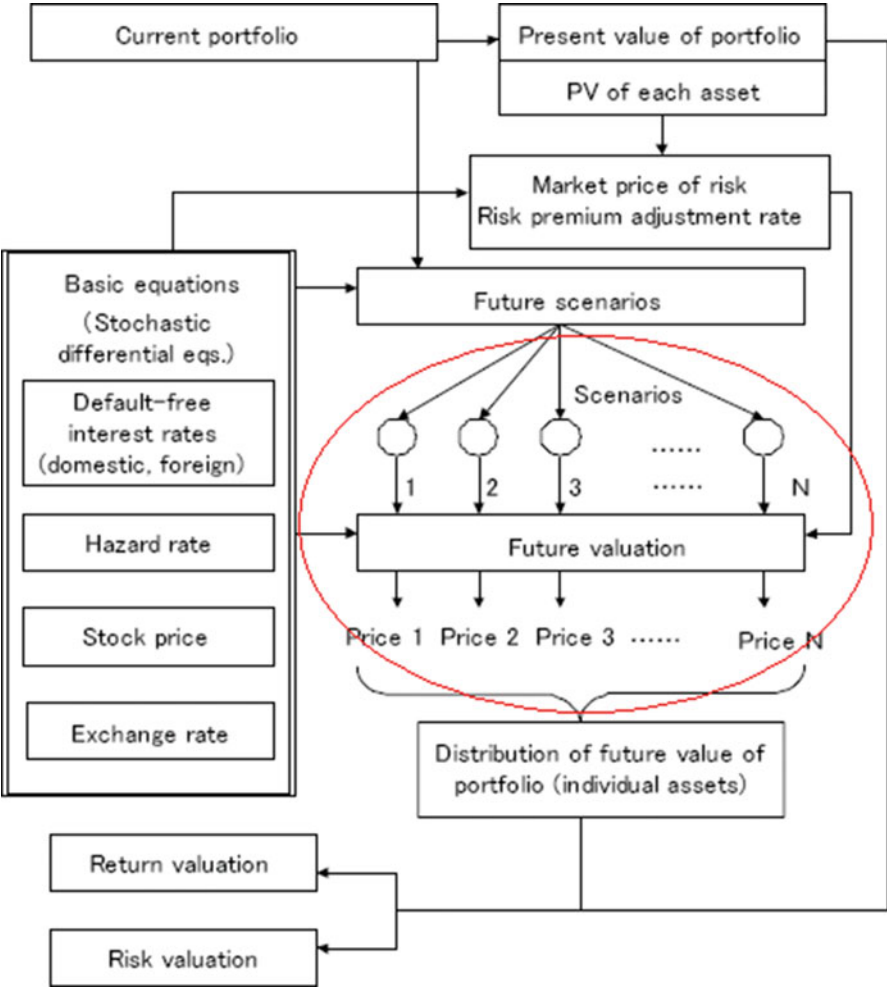


Fig. 4 Architecture of integrated risk management system

Use-case 3 – Valuing an option-based product. The use-cases were chosen to provide examples of three distinct patterns of use – aggregation, standard selection, and multiple selection. The architecture (see Fig. 5) is bound specifically with the use-cases. The advantage for grid in this case is that it can be easily tailored into specific user need to integrate different applications, which is a crucial strength of using grid.

80.2.2 Data-Intensive IT Systems

Grid in financial services from the perspective of web services toward financial services industry. The perspective is more on the transactional side. Once the

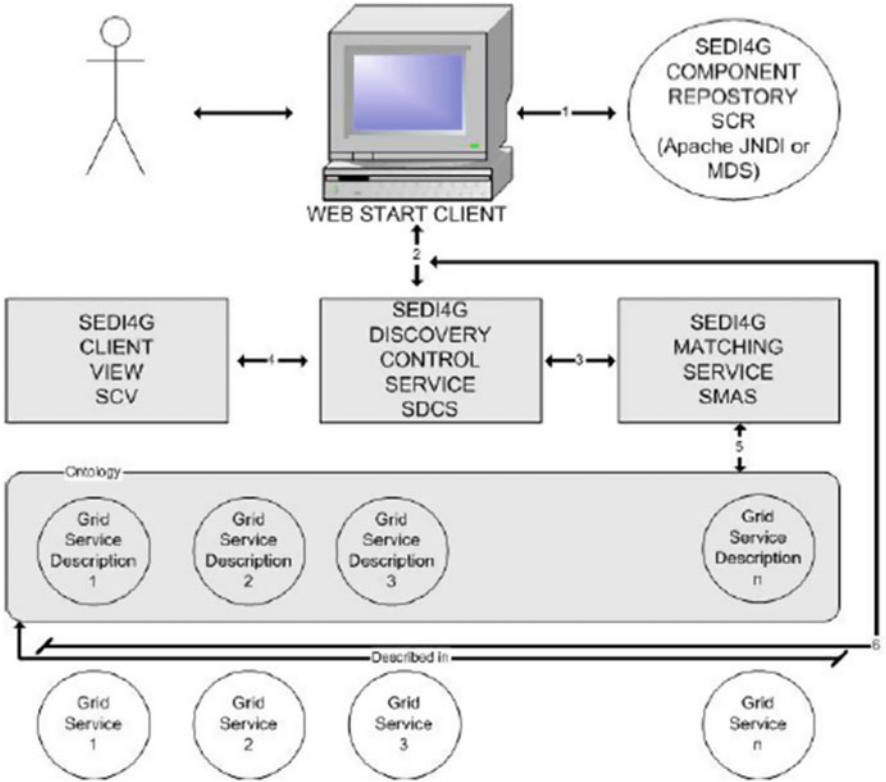


Fig. 5 The Semantic discovery for grid services architecture

bottleneck of compute cycle is solved, the data-centric nature plays the key role again. The knowledge flows back to the customized business; logic should provide the best path for users to access the live data of interest. There is no strong focus of development on this data-intensive grid system. Even in FinGrid, which claims in streaming live data for real-time analysis, the data issue remains part of compute grids. However, the need for dynamic data management is obvious as mentioned in Amad et al. (2004).

RBNB DataTurbine was used to support global environmental observatory network, which involves linking with tens of thousands of sensors and is able to obtain the observed data online. It meets grid/cyberinfrastructure (CI) requirements with regard to data acquisition, instrument management, and state-of-health monitoring including reliable data capture and transport, persistent monitoring of numerous data channels, automated processing, event detection and analysis, integration across heterogeneous resources and systems, real-time tasking and remote operations, and secure access to system resources. To that end, streaming data middleware provides the framework for application development and integration.

Use cases of RBNB DataTurbine include adaptive sampling rates, failure detection and correction, quality assurance, and simple observation (see Tilak et al. 2007). Real-time data access can be used to generate interest and buy-in from various stakeholders. Real-time streaming data is a natural model for many applications in observing systems, in particular event detection and pattern recognition. Many of these applications involve filters over data values, or more generally, functions over sliding temporal windows. The RBNB DataTurbine middleware provides a modular, scalable, robust environment while providing security, configuration management, routing, and data-archival services. The RBNB DataTurbine system acts as an intermediary between dissimilar data monitoring and analysis devices and applications. As shown in Fig. 6, a modular architecture is used, in which a source or “feeder” program is a Java application that acquires data from an external live data source and feeds it into the RBNB server. Additional modules display and manipulate data fetched from the RBNB server. This allows flexible configuration where RBNB serves as a coupling between relatively simple and “single purpose” suppliers of data and consumers of data, both of which are presented a logical grouping of physical data sources. RBNB supports the modular addition of new sources and sinks with a clear separation of design, coding, and testing. From the perspective of distributed systems, the RBNB DataTurbine is a “black box” from which applications and devices send data and receive data. RBNB DataTurbine handles all data management operations between data sources and sinks, including reliable transport, routing, scheduling, and security. RBNB accomplishes this through the innovative use of memory and file-based ring buffers combined with flexible network objects. Ring buffers are a programmer-configurable mixture of memory and disk, allowing system tuning to meet application-dependent data management requirements. Network bus elements perform data stream multiplexing and routing. These elements combine to support seamless real-time data archiving and distribution over existing local and wide area networks. Ring buffers also connect directly to client applications to provide streaming-related services including data stream subscription, capture, rewind, and replay. This presents clients with a simple, uniform interface to real-time and historical (playback) data.

80.2.3 Cloud-Enabled Big Data and Machine Learning IT Systems

80.2.3.1 Big Data and MapReduce Systems

Big Data emerged as a leading IT technology development in the early 2010s. The main driving force behind can be considered as the disruption of smart phones and the broad use of Internet. The definition of big data is formally given by NIST (Chang and Grady 2015) which consists of extensive datasets, primarily in the characteristics of volume, variety, velocity, and/or variability, that require a scalable architecture for efficient storage, manipulation, and analysis. The most popular big data tool that has been widely adopted is Hadoop (Cutting 2018). It adopts similar map-reduce algorithm that Google search used and is made to be an open source so that the general public can be benefited. The ecosystem of Hadoop has been

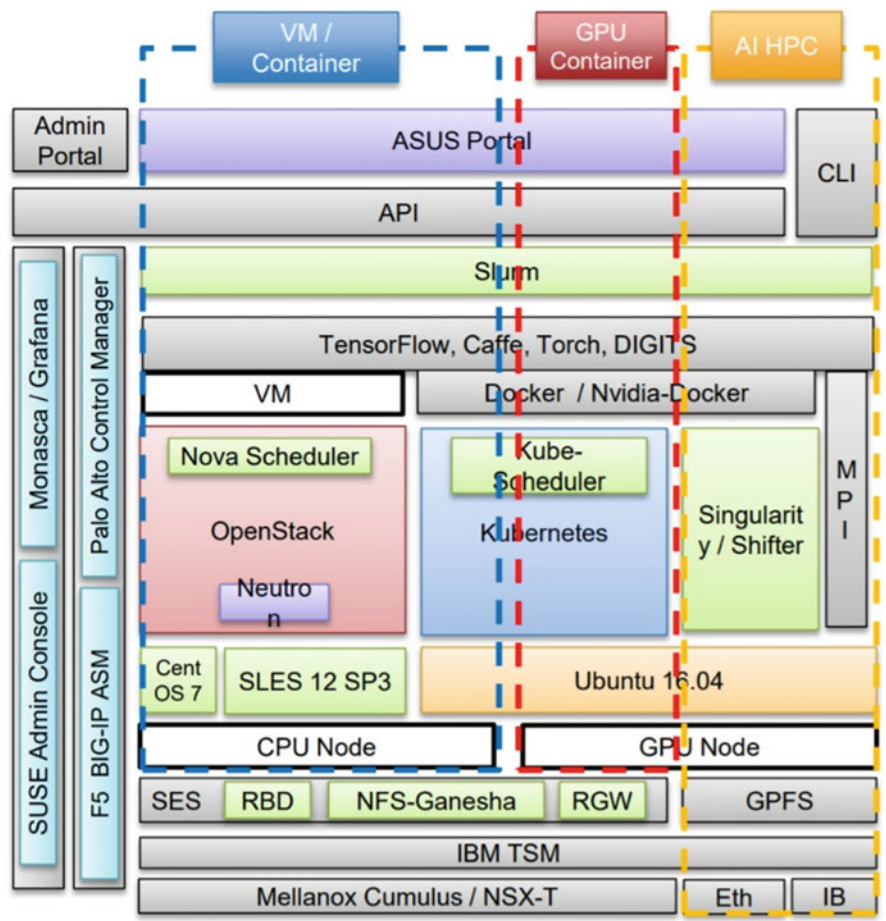


Fig. 6 Taiwan system architecture

booming since then. The idea is to cluster distributed storages into a single virtual one. Data is so-called sharded into pieces with multiple copies, usually three copies, scattering in the clustered distributed storages. To take advantage of the architecture, one needs to use map-reduce algorithm. Google search uses page ranking algorithm, which is also one kind of map-reduce realization (Bhawiyuga and Kirana 2016).

80.2.3.2 Cloud-Enabled Hybrid Systems for Big Data and Machine Learning

Alongside with the development of IT systems for big data, there has been a crucial development in coprocessors that are specifically used to accelerate the performance of the original systems. One of the coprocessors in mainstream is the graphic process unit (GPU), which is originally designed to accelerate computer graphics in usual personal computers. The performance of GPU in graphics has been enhanced not

only as exceedingly powerful but also affordable for layman due to its mass market in game industry. NVIDIA is one of the leaderships and the most well received in GPU solutions. It was gradually introduced to and developed for HPC in the 2010s, which leads to the current hybrid system design.

The hybrid system design is currently in mainstream. In the newest TOP 500 list, which ranks the most powerful supercomputers in the world, published in December 2020, six out of the top ten supercomputers adopt NVIDIA acceleration, such as its Tesla V100 and A100 GPUs. For example, Summit, an IBM-built system at the Oak Ridge National Laboratory (ORNL), remains the fastest system in the USA with a performance of 148.8 petaflops. Summit has 4356 nodes, each one housing two 22-core Power9 CPUs and six NVIDIA Tesla V100 GPUs. NVIDIA also introduced GPGPU (general purpose GPU) embedded into supercomputers, which can be used for floating point numerical analysis directly without sacrificing performance significantly.

One of the state-of-the-art systems is Taiwan 2, which is used to conduct our experiments. Taiwan 2 system consists of 252 compute nodes. Each node contains two Intel Gold Xeon CPUs, eight NVIDIA Tesla V100 GPUs, and 768GB memory with a total performance of nine petaflops. OpenStack and Kubernetes are used to virtualize system resources of HPC and OpenShift for pipeline services specifically for data analytics. The system exemplifies how such convergence can be achieved. Due to its capability, the system is also known as artificial intelligence (AI) supercomputer. The general architecture of Taiwan system is shown in Fig. 6.

Finally, data privacy and security are fundamental to big data and machine learning not only specifically to financial services. Most of the current development is based on algorithms with distributed IT systems, e.g., federated learning (Yang et al. 2019; Xian et al. 2020) and differential privacy (Gandhi and Jayanti 2020). There are system level studies specifically on cloud-based hybrid systems, which are beyond the scope of this chapter and worth a separate entry for introduction.

80.3 Distributed and Parallel Financial Simulation

In the previous sections, we addressed issues of incorporating IT technology for financial competitiveness and derive that the core lies on the performance of the IT platform, provided the competitors in the market have similar capacity and are equally informed. Grid technology, as the leading IT development in high-performance computing, is introduced as the cutting edge IT platform to meet our goal. Many companies have adopted similar technology of grids with success as mentioned in Sect. 80.1. There are also increasing research interests, which result in the work discussed in Sect. 80.2. Better performance, however, cannot be achieved by merely using a single architecture as observed in the cases of Sect. 80.2. The architecture obviously has to be specifically chosen for the analysis of interest. Simultaneously, the analysis procedures have to be tailored into the chosen architecture for performance fine tune.

In this section, we will introduce and discuss analysis procedures of financial simulation and how to tailor the analysis procedures into grid architectures by distribution and parallelism. The popular calculations for option pricing and for value at risk (VaR) in trading practice are used to serve the purpose. The calculation is based on Monte Carlo simulation, which is chosen not only because it is a well-received approach due to the absence of straightforward closed form solutions for many financial models, but also a numerical method intrinsically suited to mass distribution and mass parallelism. The success of Monte Carlo simulation lies on the quality of random number generator, which will be discussed in detail at the end of the section.

80.3.1 Financial Simulation

There is a wide variety of sophisticated financial models developed, to name a few, ranging from analysis in time series, fractals, nonlinear dynamics, and agent-based modeling to applications in optional pricing, portfolio management, market risk measure, etc. (Schmidt 2005), in which option pricing and VaR calculations of market risk measure can be considered crucial and one of the most practiced activities in market trading.

80.3.1.1 Option Pricing

An option is an agreement between two parties to buy or sell an asset at a certain time in the future for a certain price. There are two types of options:

Call Option: A call option is a contract that gives the right to its holder (i.e., buyer) without creating an obligation, to buy a prespecified underlying asset at a predetermined price. Usually, this right is created for a specific time period, e.g., 6 months, or more. If the option can be exercised only at its expiration (i.e., the underlying asset can be purchased only at the end of the life of the option), the option is referred to as a European style call option (or European call). If it can be exercised on any date before its maturity, the option is referred to as an American style call option (or American call).

Put option: A put option is a contract that gives its holder the right without creating the obligation, to sell a prespecified underlying asset at a predetermined price. If the option can be exercised only at its expiration (i.e., the underlying asset can be sold only at the end of the life of the option), the option is referred to as a European style put option (or European put). If it can be exercised any date before its maturity, the option is referred to as an American style put option (or American put).

To price options in computational finance, we use the following notation: K is the strike price; T is the time to maturity of the option; S_t is the stock price at time t ; r is the risk-free interest rate; μ is the drift rate of the underlying asset (a measure of the average rate of growth of the asset price); σ is the volatility of the stock; and

V denotes the option value. Here is an example to illustrate the concept of option pricing. Suppose an investor enters into a call option contract to buy a stock at price K after 3 months. After 3 months, the stock price is S_t . If $S_t > K$, then one can exercise one's option by buying the stock at price K and by immediately selling in the market to make a profit of $S_t - K$. On the other hand, if $S_t \leq K$, then one should not exercise the option to buy the stock. Hence, we see a call option to buy the stock at time T at price K will get payoff $(S_T - K)^+$, where $(S_T - K)^+ \equiv \max(S_T - K, 0)$ (Schmidt 2005; Hull 2003).

80.3.1.2 Market Risk Measurement Based on VaR

Market risks are the prospect of financial losses, or gains, due to unexpected changes in market prices and rates. Evaluating the exposure to such risks is nowadays of primary concern to risk managers in financial institutions. Until the late 1980s, market risks were estimated through gap and duration analysis (interest rates), portfolio theory (securities), sensitivity analysis (derivatives), or scenarios analysis. However, all these methods either could be applied only to very specific assets or relied on subjective reasoning.

Since the early 1990s, a commonly used market risk estimation methodology has been the value at risk (VaR). A VaR measure is the highest possible loss L incurred from holding the current portfolio over a certain period of time at a given confidence level (Dowd 2002)

$$P(L > VaR) \leq 1 - c \quad (1)$$

where c is the confidence level, typically 95%, 97.5%, or 99% and P is cumulative distribution function. By convention, $L = -\Delta X(\tau)$, where $\Delta X(\tau)$ is the relative change (return) in portfolio value over the time horizon τ . Hence, large values of L correspond to large losses (or large negative returns).

The VaR figure has two important characteristics: 1) It provides a common consistent measure of risk across different positions and risk factors; and 2) it takes into account the correlations or dependencies between different risk factors. Because of its intuitive appeal and simplicity, it is no surprise that in a few years value at risk has become the standard risk measure used around the world. However, VaR has a few deficiencies, among them the nonsubadditivity – a sum of VaR's two portfolios can be smaller than the VaR of the combined portfolio. To cope with these shortcomings, Artzner et al. proposed an alternative measure that satisfies the assumptions of a coherent risk measure. The expected shortfall (ES), also called expected tail loss (ETL) or conditional VaR, is the expected value of the losses in excess of VaR:

$$ES = E(L|L > VaR) \quad (2)$$

It is interesting to note that although new to the finance industry – expected shortfall has been familiar to insurance practitioners for a long time. It is very similar to the mean excess function which is used to characterize claim size distribution (Cizek et al. 2011).

The essence of the VaR and ES computations is estimation of low quantiles in the portfolio return distributions. Hence, the performance of market risk measurement methods depends on the quality of distribution assumptions on the underlying risk factors. Many of the concepts in theoretical and empirical finance developed over the past decades, including the classical portfolio theory, the Black-Scholes-Merton option pricing model, and even the RiskMetrics variance-covariance approach to VaR, rest upon the assumption that asset returns follow a normal distribution. The assumption is not justified by real market data. Our interest is more on the calculation side. For interested readers, we refer further to Weron (2004).

80.3.2 Monte Carlo Simulation

80.3.2.1 Monte Carlo and Quasi-Monte Carlo Methods

In general, Monte Carlo (MC) and Quasi-Monte Carlo (QMC) methods are applied to estimate the integral of function $f(x)$ over $[0, 1]^d$ unit hypercube where d is the dimension of the hypercube.

$$I = \int_{[0, 1]^d} f(x) dx \quad (3)$$

In MC methods, I is estimated by evaluating $f(x)$ at N independent points randomly chosen from a uniform random distribution over $[0, 1]^d$ and then evaluating the average

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (4)$$

From the law of large numbers, $\hat{I} \rightarrow I$ as $N \rightarrow \infty$. The standard deviation is

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^N (f(x_i) - I)^2} \quad (5)$$

Therefore, the error of MC methods is proportional to $N^{-1/2}$.

QMC methods compute the above integral based on low-discrepancy (LD) sequences. The elements in a LD sequence are “uniformly” chosen from $[0, 1]^d$ rather than “randomly.” The discrepancy is a measure to evaluate the uniformity of points over $[0, 1]^d$. Let $\{q_n\}$ be a sequence in $[0, 1]^d$, and the discrepancy D_N^* of q_n is defined as follows, using Niederreiter’s notation (Niederreiter 1992):

$$D_N^*(q_n) = \sup_{B \in [0, 1]^d} \left| \frac{A(B, q_n)}{N} - v_d(B) \right| \quad (6)$$

where B is a subcube of $[0, 1]^d$ containing the origin, $A(B, q_n)$ is the number of points in q_n that fall into B , and $v_d(B)$ is the d -dimensional Lebesgue measure of B . The elements of q_n are said to be uniformly distributed if its discrepancy $D_N^* \rightarrow 0$ as $N \rightarrow \infty$. From the theory of uniform distribution sequences (Kuipers and Niederreiter 1974), the estimate of the integral using a uniformly distributed sequence $\{q_n\}$ is $\hat{I} = \frac{1}{N} \sum_{n=1}^N f(q_n)$, as $N \rightarrow \infty$ then $\hat{I} \rightarrow I$. The integration error bound is given by the Koksman-Hlawka inequality:

$$\left| I - \frac{1}{N} \sum_{n=1}^N f(q_n) \right| \leq V(f) D_N^*(q_n) \quad (7)$$

where $V(f)$ is the variation of the function in the sense of Hardy and Krause (see Kuipers and Niederreiter 1974), which is assumed to be finite.

The inequality suggests a smaller error can be obtained by using sequences with smaller discrepancy. The discrepancy of many uniformly distributed sequences satisfies $O((\log N)^d/N)$. These sequences are called lowdiscrepancy (LD) sequences (Chen et al. 2006). Inequality (7) shows that the estimates using a LD sequence satisfy the deterministic error bound $O((\log N)^d/N)$.

80.3.2.2 Monte-Carlo Simulations for Option Pricing

Under the risk-neutral measure, the price of a fairly valued European call option is the expectation of the payoff $E[e^{-rT}(S_T - K)^+]$. In order to compute the expectation, Black and Scholes (Black and Sholes 1973) modeled the stochastic process generating the price of a nondividend-paying stock as geometric Brownian motion:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \quad (8)$$

where W is a standard Wiener process, also known as Brownian motion. Under the risk-neutral measure, the drift μ is set to $\mu = r$.

To simulate the path followed by S , suppose the life of the option has been divided into n short intervals of length Δt ($\Delta t = T/n$), the updating of the stock price at $t + \Delta t$ from t is (Hull 2003):

$$S_{t+\Delta t} - S_t = r S_t \Delta t + \sigma S_t Z \sqrt{\Delta t} \quad (9)$$

where Z is a standard random variable, i.e., $Z \sim (0, 1)$. This enables the value of $S_{\Delta t}$ to be calculated from initial value S_t at time Δt , the value at time $2\Delta t$ to be calculated from $S_{\Delta t}$, and so on. Hence, a completed path for S has been constructed.

In practice, in order to avoid discretization errors, it is usual to simulate $\ln S$ rather than S . From Itô's lemma, the process followed by $\ln S$ is (Bratley and Fox 1988):

$$d \ln S = \left(r - \frac{\sigma^2}{2} \right) dt + \sigma dz \quad (10)$$

so that

$$\ln S_{t+\Delta t} - \ln S_t = \left(r - \frac{\sigma^2}{2}\right)dt + \sigma Z\sqrt{\Delta t} \quad (11)$$

or equivalently:

$$S_{t+\Delta t} = S_t \exp \left[\left(r - \frac{\sigma^2}{2}\right)dt + \sigma Z\sqrt{\Delta t} \right] \quad (12)$$

Substituting independent samples $Z_1, \dots, Z_n$ from the normal distribution into (49.12) yields independent samples $S_T^{(i)}$, $i = 1, \dots, n$, of the stock price at expiry time T . Hence, the option value is given by

$$V = \frac{1}{n} \sum_{i=1}^n V_i = \frac{1}{n} \sum_{i=1}^n e^{-rT} \max \left[S_T^{(i)} - K, 0 \right] \quad (13)$$

The QMC simulations follow the same steps as the MC simulations, except that the pseudorandom numbers are replaced by LD sequences. The basic LD sequences known in literature are Halton (1960), Sobol (1967), and Faure (1982). Niederreiter (1992) proposed a general principle of generating LD sequences. In finance, several examples have shown that the Sobol sequence is superior to others. For example, Galanti and Jung (1997) observed that “the Sobol sequence outperforms the Faure sequence, and the Faure marginally outperforms the Halton sequence.” In this research, we use Sobol sequence in our experiments. The generator used for generating the Sobol sequence comes from the modified algorithm 659 of Joe and Kuo (2003).

80.3.2.3 Monte-Carlo Bootstrap for VaR

Monte Carlo simulation is applicable with virtually any model of changes in risk factors and any mechanism for determining a portfolio’s value in each market scenario. But revaluing a portfolio in each scenario can present a substantial computational burden, and this motivates research into ways of improving the efficiency of Monte Carlo methods for VaR.

The bootstrap (Efron 1981; Efron and Tibshirani 1986) is a simple and straightforward method for calculating approximated biases, standard deviations, confidence intervals, and so forth, in almost any nonparametric estimation problem. Method is a key word here, since little is known about the bootstrap’s theoretical basis, except that (a) it is closely related to the jackknife in statistic inferring; (b) under reasonable condition, it gives asymptotically correct results; and (c) for some simple problems which can be analyzed completely, for example, ordinary linear regression, the bootstrap automatically produces standard solutions.

The bootstrap method is straightforward. Suppose we observe returns $X_i = x_i$, $i = 1, 2, \dots, n$, where the X_i are independent and identically distributed (iid) according to some unknown probability distribution F . The X_i may be real valued, two-dimensional, or take values in a more complicated space. A given parameter $\theta(F)$, perhaps the mean,

median, correlation, and so forth, is to be estimated, and we agree to use the estimate $\hat{\theta} = \theta(\hat{F})$, where $\hat{F}$ is the empirical distribution function putting mass $1/n$ at each observed value x_i . We wish to assign some measure of accuracy to $\hat{\theta}$.

Let $\sigma(F)$ be some measure of accuracy that we would use if F were known, for example, $\sigma(F) = \text{SD}_F(\hat{\theta})$, the standard deviation of $\hat{\theta}$ when $X_1, X_2, \dots, X_n \sim F^{(\text{id})}$.

The bootstrap estimate of accuracy is $\hat{\sigma} = \sigma(\hat{F})$ which is the nonparametric maximum likelihood estimate of $\sigma(F)$. In order to calculate $\hat{\sigma}$, it is usually necessary to employ numerical methods. (a) A bootstrap sample $X_1^*, X_2^*, \dots, X_n^*$ is drawn from $\hat{F}$, in which each X_i^* independently takes value x_j with probability $1/n, j = 1, 2, \dots, n$. In other words, $X_1^*, X_2^*, \dots, X_n^*$ is an independent sample of size n drawn with replacement from the set of observations $\{x_1, x_2, \dots, x_n\}$. (b) This gives a bootstrap empirical distribution function $\hat{F}^*$, the empirical distribution of the n values $X_1^*, X_2^*, \dots, X_n^*$, and a corresponding bootstrap value $\hat{\theta}^* = \theta(\hat{F}^*)$. (c) Steps (a) and (b) are repeated, independently, a large number of times, say N , giving bootstrap values $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*N}$. (d) The value of $\hat{\sigma}$ is approximated, in the case where $\sigma(F)$ is the standard deviation by the sample standard deviation of the $\hat{\theta}^*$ values, where

$$\hat{\mu} = \frac{\sum_{j=1}^n \hat{\theta}^{*j}}{N} \quad (14)$$

and

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^n (\hat{\theta}^{*j} - \hat{\mu})^2}{N - 1} \quad (15)$$

80.4 Case Study and Discussions

80.4.1 Case Study

80.4.1.1 Asian Options and Rainbow Options

To demonstrate what the grid computing can contribute to the financial service in a significant manner, two kinds of popular options, Asian options and rainbow options, are chosen for Monte Carlo pricing model. Asian options have payoffs that depend on the average price of the underlying asset such as stocks, commodities, or financial indices. However, there is no exact closed-form formula existing for these popular options. Rainbow options, also known as basket options, are referred to an entire class of options which consist of more than one underlying asset. Rainbow options are usually calls or puts on the best or worst of the underlying assets, or options which pay the best or worst of the assets. They are excellent tools

for hedging risk of multiple assets. The rainbow options are therefore used for our bootstrap calculations of VaR.

80.4.1.2 Parallelization, Distribution, and Message Passing Interface (MPI)

MPI is a library specification for message passing, proposed and developed as a standard by a broadly based committee of vendors, developers, and users (Snir et al. 1996). MPI was designed for high performance on both massively parallel machines and on distributed clusters. The MPI standard is nowadays widely accepted and used in the community of high-performance computing.

The basic MPI functions are point-to-point pair-wise message passing for send and for receive. Collective communications are also provided for ease of use as well as better performance. These communication methods, when used in supercomputers, do facilitate the parallelization of numerical methods that require both heavy compute cycles and stronger dependency between parallel processes.

Recent development of supercomputer, affected by the popularity of cluster computing in PCs, tends to be designed hierarchically scalable. Further extension of clusters of supercomputers can be regarded as an initial concept of grids (see Sect. 80.2). The use of MPI is straightforward in this kind of hardware architecture and interlink of networks. There is always an obvious physical limitation in this architecture, which is also proportional to the limitation of the investment of governmental research funding. People tend to use mass distribution of computers, mostly PCs, which are linked loosely in Internet cloud. The terminology cloud is often used in networking community to show that in Internet there is no specific network path from one computer to another. MPI working in such an environment is expected to be inefficient and unstable, e.g., high network latency-induced packet lost in long distance real-time communication. In the following sections, three specific platforms, including local clusters, geographically distributed large clusters, and PC grids with tens of thousands of PCs connected in the cloud, will be used for the financial calculations to demonstrate benefits in using grids.

80.4.1.3 Empirical Study for Data Grid System

In order to demonstrate the usefulness of grid system, in particular in data-intensive application, the real market data are used, including daily from iShares (Morgan Stanley Capital International), MSCI Taiwan Index (ETF), and Taiwan Stock Exchange Center (TSEC) weighted index, extracted specifically from May 31, 2005, to May 31, 2008, and 30 days tick-by-tick trading data from Taiwan Futures Exchange Center (TAIFEX).

80.4.2 System Platform Tests

The various grid platforms are carefully chosen to demonstrate the performance issue in finance services, which include a big data Hadoop Cluster, Taiwania 2 cloud-enabled AI supercomputer, a small diskless remote boot Linux (DBRL)

PC clusters, large-scale and geographically widely distributed test-bed the Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) compute grid, and a densely distributed at-home style PC grid, which resembles the clouding computing.

80.4.2.1 Artificial Intelligence-Based High-Performance Computing Platform

As introduction in Sect. 80.2, an Apache Hadoop system is employed to tackle and demonstrate big data solutions. The example given here is for fraud detection based on cyclic trading pattern. The original dataset is unified invoices from the Financial Information Agency of Taiwan, collected over a 3-year period from 2010 to 2012 consisting 378,373,684 records. The number of companies involved in the trading process of interest is 1,385,870, excluding the listed companies in stock market. The company is deidentified, and all other information are removed for this study, but the temporal sequence is kept in order. The dataset is used to search long path cycles using brutal force approach.

In the current case, it is more to do with the volume of data. Practically, the raw data is only approximately a few gigabytes. During the calculation, the intermittent data size can reach almost petabytes, which is almost ten orders of magnitude. The calculation is performed on Apache Hadoop in a supercomputer of 4096 cores (16 core CPU), 16.38 TB memory, and 1.5 PB HDFS (Hadoop File System). The platform based on Hadoop allows us to handle a single file in petascale in this case, which is exactly what big data is designed for. For technical details and results, we refer to Lin et al. (2020).

Similar big data cluster can be built in Taiwan 2 system using virtual machines, which is versatile not only for big data but also for general machine learning. The software architecture of Taiwan 2 is shown in Fig. 7. From the cloud terminology, it ranges from infrastructure as a service (IaaS), platform as a service (PaaS), to software as a service (SaaS). The machine learning from various neural networks, such as deep neural networks, recursive neural networks, and reinforcement neural networks, has been developed way beyond its original scope achieving great successes in computer vision, speech and audio, and national language processing. The technology is now well adopted into a broad range of financial services, which use cases include (Financial Stability Board Report 2017):

- To assess credit quality, to price and market insurance contracts, and to automate client interaction by financial institutes and vendors
- To optimizing scarce capital as well as back-testing models and analyzing the market impact of trading large positions by financial institutes
- To find signals for higher (and uncorrelated) returns and optimize trading execution by hedge funds, broker-dealers, and other firms
- For regulatory compliance, surveillance, data quality assessment, and fraud detection by both public and private sector institutes

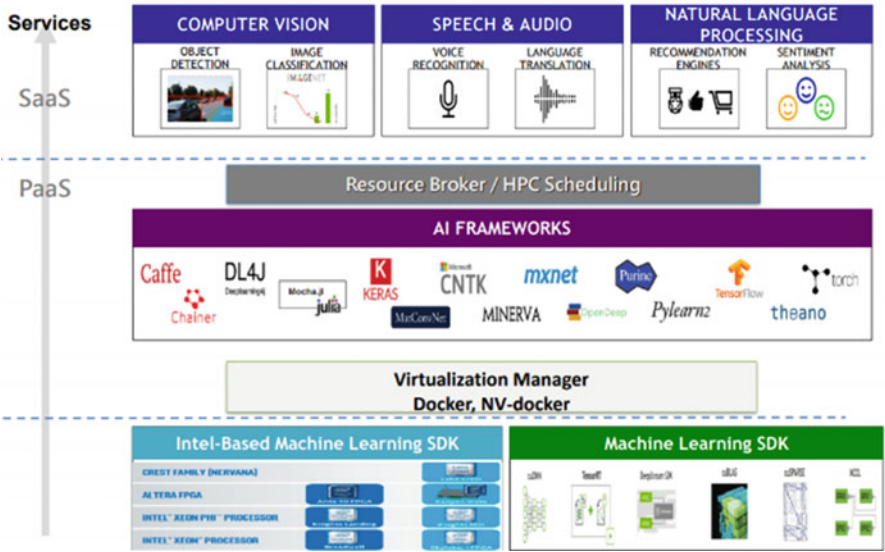


Fig. 7 The software stack of Taiwania 2 system

Taiwania 2 provides compute capability for traditional modeling based on Monte-Carlo Simulation and then extends to big data and machine learning. In our example, a LSTM- (Hochreiter and Schmidhuber 1997) recursive neural network and reinforcement neural network are used to train models to predict trends of foreign exchange rate and to derive a trading strategy for foreign exchange rate. The data is online streamed from Yahoo Finance for both training/testing and predicting in real time. Taiwania 2 system effectively builds models and predicts as expected using the AI frameworks from Fig. 7.

80.4.2.2 Diskless Remote Boot Linux (DRBL) Cluster

DRBL is an in-house program of National Center for High-Performance Computing (NCHC) and was an original software product developed by Steven Shiau and his group under the auspice of National Knowledge Innovation Grid (KING) of Taiwan. It was developed initially as a centralized system management tool aiming at small and median size PC clusters. It is nowadays recognized internationally as one of the most advanced mass backup solutions. Here, DRBL is used as an alternative scavenger for compute cycles of spare clusters. When needed, it converts systems of PC clusters into an aggregated and homogenous Linux system, simultaneously with a mass backup of the original systems, and recovery backup of the original systems once the need is satisfied. The most popular use is to convert a PC classroom into a compute Linux cluster. In such a case, compute cycles of the clusters can be fully exploited. In a grid environment, this is a perfect case to resource scale-up when in contingent need and once the situation-relieved resources will be released correspondingly. Such a dynamic feature can be beneficial for the financial services.

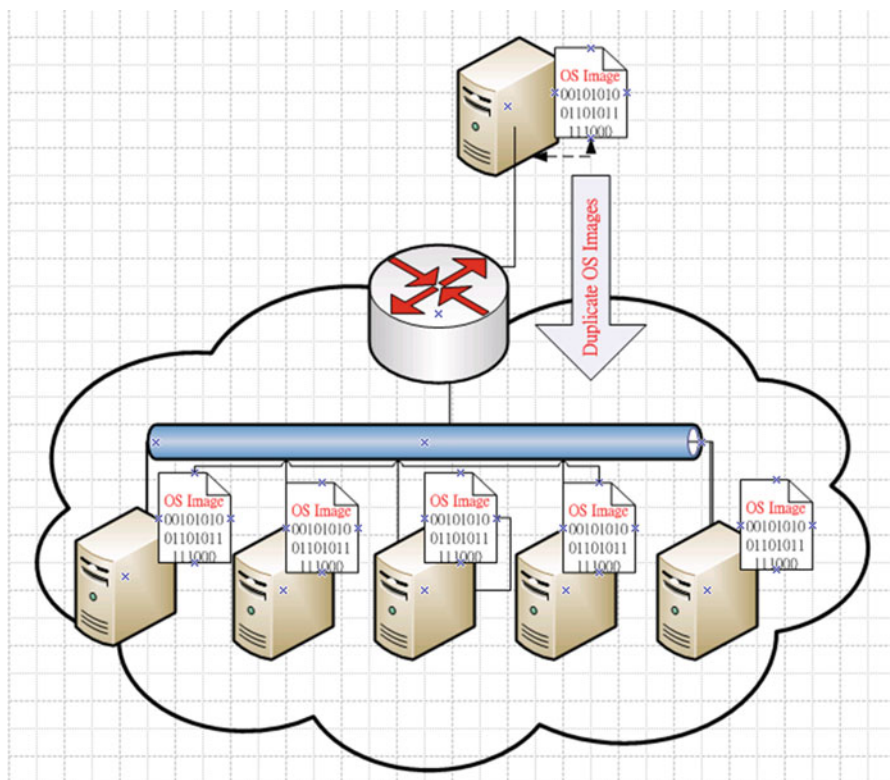


Fig. 8 The schematic of DRBL system

The schematic of DRBL system can be shown in Fig. 8, where DRBL duplicates image files of an operational system, e.g., Linux kernel, via network to the clients. The clients' original operational systems are not used. The clients are therefore temporarily turned into dedicated compute resources, which also provide additional security to financial data.

The test case here involves Monte Carlo simulations on Asian option pricing, on rainbow option pricing, and on a bootstrap VaR calculation, respectively (see section “[Distributed and Parallel Financial Simulation](#)”). The market parameters are given as risk-free interest rate $r = 0.1$, underlying asset spot price $S = 100$, strike price $X = 100$, duration to maturity $T = 125$, and volatility $\sigma = 0.3$. For the rainbow options, a linear weighted combination of four underlying assets is assumed with underlying prices of $S_i = 100, 110, 120, 130$ and the corresponding weightings of volatility $\sigma_i = 0.3, 0.4, 0.5, 0.6$. The correlation matrix is taken to be

$$\rho_{ij} = \begin{pmatrix} 0.5 & 0.4 & 0.5 \\ 0.4 & 0.3 & 0.4 \\ 0.5 & 0.4 & 0.6 \end{pmatrix}.$$

Table 1 Comparison of performance between DRBL-based PC platform with 32 nodes, FORMOSA II of NCHC with a batch job of 32 nodes and IBM cluster 1350 with a batch job of 32 nodes. The PCs are 20 XEON 2.6 GHz and 4GB RAM

	DRBL cluster	FORMOSA II	IBM cluster 1350
Asian option pricing (AOP)	81	49	24
Rainbow option pricing (ROP)	329	197	98
VaR calculation (based on ROP w/ bootstrap)	322	194	97

Unit: microsecs per Monte Carlo path. Averaged from 1000,000 Monte Carlo paths

Table 2 Comparison of speedup ratios based on the calculations in Table 1

	DRBL cluster	FORMOSA II	IBM cluster 13 50
Asian option pricing (AOP)	28.68	29.46	30.16
Rainbow option pricing (ROP)	27.31	28.59	29.34
VaR calculation (based on ROP w/ bootstrap)	27.20	28.10	29.10

Unit: Speedup ratio: CPU (nonparallel single node)/CPU (parallel single node)

The calculation of the VaR uses the same four-dimensional rainbow options with additional expectation of return 0.07, 0.08, 0.09, and 0.10 respectively. They are calculated in DRBL cluster as well as benchmarked with two cluster-based supercomputers in NCHC. The results are shown in Tables 1 and 2.

In Table 1, instead of giving a total wall clock time of the calculation with some given numbers of Monte Carlo simulations or paths, a more useful averaged single Monte Carlo simulation based on 1,000,000 simulations is used to demonstrate the performance when different system architectures are used. The results show that the traditional big irons, i.e., supercomputers, still outperform the cluster. Yet, considering there are no extra costs invested in the computing resources and still compute cycle in a sizable manner is obtained, the approach is appealing to be further developed in to a fully operational production system.

80.4.2.3 Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) Grid

The Pacific Rim Applications and Grid Middleware Assembly (PRAGMA), founded in 2002, is an open international organization that focuses on a variety of practical issues of building international scientific collaborations in a number of application areas. PRAGMA grid is established by the resources and data-computing working group as a global grid testbed for benchmarking the interoperability of grid middleware and the usability and productivity of grids. The PRAGMA grid consists of physical resources as well as system administration supports from 29 institutions across 5 continents and 14 countries. It is an instantiation of a useful, interoperable, and consistently available grid system that is neither dictated by the needs of a single science domain, nor funded by a single national agency. It does not have uniform

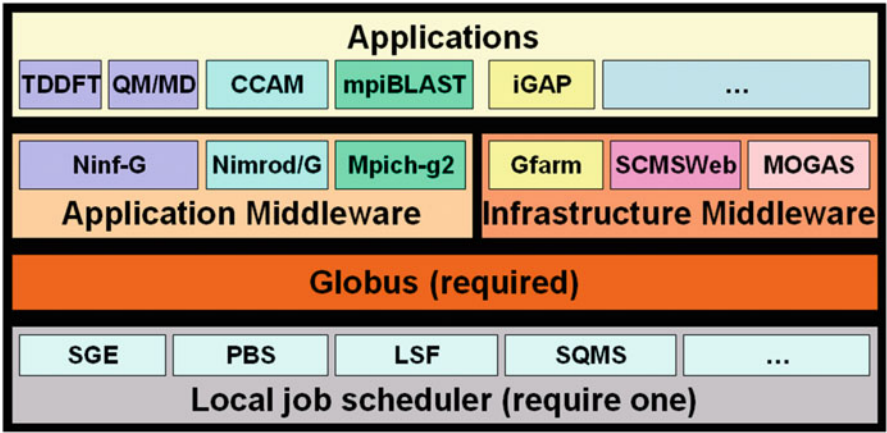


Fig. 9 Software stack developed in the PRAGMA grid

infrastructure management yet is robust and supports a wide range of scientific applications. The software stack of the system is shown in Fig. 9.

The PRAGMA grid successfully tackles the issues of distance and time zone differences among sites, lack of infrastructure tools for heterogeneous global grid, nonuniform system and network environments, and diverse application requirements. For more technical details both in theory and practice, we refer to Abramson et al. (2006).

The similar test case in Sect. 80.4 was followed, but the platform was extended with a collection of clusters across institute boundaries. The common job submission is executed via a homogeneous middleware Globus Toolkit. We demonstrate the usefulness of the platform by grouping compute resources across national boundaries and still achieve good performance. The results are shown in Tables 3 and 4.

80.4.2.4 At-Home Style PC Grid

With the continued penetration of personal computers and the remarkable improvement of CPU processing speed, 80–90% of most PCs’ processing power is untapped, according to a study. This does not mean that many PCs remain turned off, but that the capacity of the CPU, the brain of the PC, is not fully utilized. In case the CPU is more extensively used when a task requiring an enormous number of operations, such as three-dimensional graphical processing, is assigned, it sits idle most of the time during word processing and Internet browsing because CPU processing speeds are much faster than the speeds of input from the keyboard or the communication line.

This fact led to the idea of virtually gathering the power of idle CPUs to use as a computer resource. In other words, this means networking numerous computers to make them work like a single high-performance computer, and assigning complex processing tasks to it. The assigned task will be divided into a myriad of small tasks and allocated to individual computers on the grid-like network. Even when

Table 3 Comparison of performance between group A, which consists of 13 nodes from NCHC and 15 from UCSD, and group B, which consists of 122 nodes collectively from UCSD, AIST, NCHC, and Osaka University. The details of resources are referred to (<http://pragma-goc.rocksclusters.org/pragma-doc/resources.html>)

	Group A	Group B
Asian option pricing (AOP)	68	56
Rainbow option pricing (ROP)	278	234
VaR calculation (ROP w/bootstrap)	271	229

Unit: secs/per Monte Carlo path. Averaging from 1000,000 Monte Carlo paths

Table 4 Comparison of speedup ratios based on the calculations in Table 3

	Group A	Group B
Asian option pricing (AOP)	25.24	107.58
Rainbow option pricing (ROP)	25.81	110.34
VaR calculation (ROP w/bootstrap)	25.63	110.09

Unit: secs/per Monte Carlo path. CPU (nonparallel single node)/CPU (parallel single node)

increasingly faster CPUs, the power of PCs is not comparable to those of supercomputers, but in a networked environment where individual PCs simply process complex task in parallel, PCs can deliver surprisingly high performance. This is the core concept of CP grid computing (Fig. 10), and it came into a reality several years ago (Table 5).

The commoditization and the increased processing speed of PCs leads the growth of idle CPU power. This will facilitate the construction of PC grid computing system along with the improvement of the communication environment by broadband connectivity (Chen et al. 2006). In practice, a PC grid platform Korea@Home (or K@H) is used in our study. Its architecture is shown in Fig. 10. It is based on MS Windows. Asian option pricing is used to demonstrate the performance of the system, in which the number of iterations is taken to be 1,000,000. The duration to maturity is further divided by 10,000 periods, which push the system to run on the mass parallel system of K@H. To tackle this scale, or even larger scale for all kinds of possible scenarios in real trading practice, an off-line-distributed and parallel approach is adopted. The total number of Monte Carlo simulation is $1,000,000 \times 10,000$. It was divided into 10,000 jobs, and each job consists of 1,000,000 Monte Carlo simulations. The market parameters are given as above.

The results demonstrated here are not as good as expected (see Table 1). It shows that the speedup ratio is only 3.79. In this test, K@H further divides the jobs into ten groups. Each group was sent and run in a sequential fashion, which causes the low speedup ratio. However, if one looks into the executed CPU time for each job, our assumption is still valid. We simulated the result in a small cluster with the same scenario and Monte Carlo paths. The result shows 90% speedup can be easily achieved.

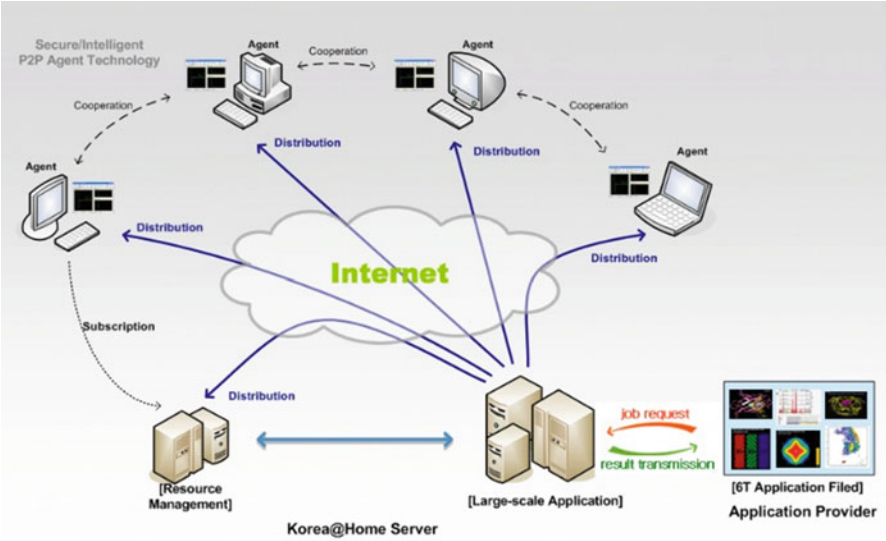


Fig. 10 The architecture of Korea@Home, a specific @Home style PC grid used in our case study (Jun-Weon Yoon 2008)

Table 5 Summary of the case for the PC grid calculations

Asian option pricing	Statistics
Number of Monte Carlo path	1,000,000 x 10,000
The running period (1) (Wall clock time)	28 h 51 m (104,911 s)
Number of jobs	10,000
CPU time per job	28 ~ 30s
Total running time (2) (Wall clock time)	4 days 14 h 31 m (397,919 s)
Speedup ratio $\frac{(2)}{(1)}$	3.79

80.4.2.5 RBNB Data Grid

RBNB DataTurbine in market data streaming is implemented here (see Fig. 11 for real-data plot in collaboration with Ebbe et al. 2007), in which the real data from iShares MSCI Taiwan index (ETF) and TSEC-weighted index from May 31, 2005, to May 31, 2008, and 30 days tick-by-tick trading data from Taiwan Futures Exchange Center (TAIFEX) are used and open in different file-based online channels from the RBNB DataTurbine. The purpose is to dynamically manage the high-frequency market data and connect the data with analysis applications on the fly. The system implemented up to date is for large-scale dynamic and static data management.

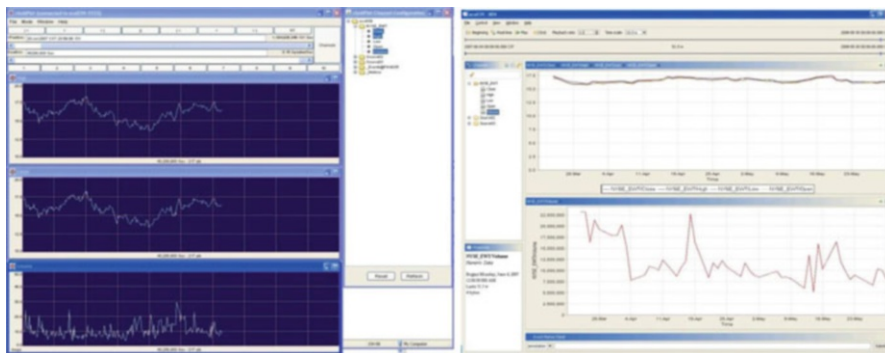


Fig. 11 RBNB DataTurbine streaming open for data channels of iShares MSCI Taiwan index (ETF) and TSEC-weighted index from May 31, 2005, to May 31, 2008, and 30 days tick-by-tick trading data from Taiwan Futures Exchange center (TAIFEX)

80.5 Conclusion

Securities trading is one of the few business activities where a few seconds processing delay can cost a company big fortune. The growing competition in the market exacerbates the situation and pushes further toward instantaneous trading even in a split second. The key lies on the performance of the underlying information system. Following the computing evolution in financial services, it was a centralized process to begin with and gradually decentralized into a distribution of actual application logic across service networks. Financial services have the tradition of doing most of its heavy lifting financial analysis in overnight batch cycles. However, in securities trading, it cannot satisfy the need due to its ad hoc nature and requirement of immediate response. A new computing paradigm, grid computing, aiming at virtualizing scale-up distributed computing resources, is well suited to the challenge posed by the capital market practices.

In this chapter, we revisit the theoretical background of how performance will affect the market competition. The core concept lies on information asymmetry. Due to the advance of IT, even in a split second it will be a matter of winning or losing in real market practice. After establishing the motivation, we reviewed recent grid development specifically used for finance service. Monte Carlo simulations are chosen not only because of their popularity in real world, but also because of their nature of so-called “fine grain” or mass parallelism approach. The success of Monte Carlo simulations lies on better random number generators. The well-recognized Sobol sequences as a quasi-random number generator are carefully studied to ensure the quality of Monte Carlo simulations when employed for mass parallelism. Then some popular basic option-pricing models, collectively Asian option pricing, rainbow option pricing, and VaR calculation with constant market parameters, are introduced as drivers to introduce more details of grids for better finance service. Finally, we test various grid platforms, based on the methodology of mass

parallelism and mass distribution, with the drivers. The real market data are also used, but at this stage they are only used to demonstrate the dynamic data management, in which grids can offer better.

Our future work will migrate grid platforms to cloud platform with HPC capability and extend them to broader use of big data and machine learning applications for more general financial services. Along the current track of this development, we will also develop models based on machine learning for option pricing and risk management with tick-by-tick daily market information.

References

- Abramson, D., et al. 2006. Deploying scientific applications to the PRAGMA Grid Testbed: Strategies and lessons. In *Sixth IEEE international symposium on cluster computing and the grid*, 241–248.
- Ahmad, K., Gillam, L., and D. Cheng. 2005. Textual and quantitative analysis: Towards a new, e-mediated social science. In *Proceedings of the 1st international conference on e-social science*, Manchester, 22–24.
- Almond, J., and D. Snelling. 1999. UNICORE: Uniform access to supercomputing as an element of electronic commerce. *Future Generation Computer Systems* 15 (5): 539–548.
- Amad, K., et al. 2004. Financial information grid –an ESRC e-Social science pilot. In *Proceedings of the third UK e-science programme all-hands meeting*, Nottingham.
- Atkins, D.E., et al. 2003. Revolutionizing science and engineering through cyberinfrastructure. In *Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyber-infrastructure*. National Science Foundation.
- Bell, D., and A.S. Ludwig. 2005. Grid service discovery in the financial markets sector. In *7th international conference on information technology interfaces*, Cavtat, Croatia.
- Bhawiyuga, A., and A.P. Kirana. 2016. Implementation of page rank algorithm in Hadoop MapReduce framework. *ISITIA, IEEE Xplore*.
- Black, F., and M. Sholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–654.
- Bratley, P., and B.L. Fox. 1988. Algorithm 659: Implementing Sobol's Quasirandom sequence generator. *ACM Transactions on Mathematical Software* 14: 88–100.
- Bughin, J., Chui, M., and J. Manyika. 2010. Cloud, Big Data and smart assets: Ten Tech-enabled business trends to watch. *McKinsey Quarterly* August, 7–8.
- Chang, W.L., and N. Grady. 2015. NIST Big Data interoperability framework. *Big Data Definitions, Special Publication (NIST SP) – 1500-1*.
- Chang, F., Dean, J., Ghemawat, S., Hsieh, W.C., Wallach, D.A., Burrows, M., Chandra, T., Fikes, A., and R.E. Gruber. 2006. BigTable: A distributed storage system for structured data. In *OSDI '06: Proceedings of the 7th USENIX symposium on operating systems design and implementation*, 15–15.
- Chen, G., Thulasiraman P., and R.K. Thulasiram. 2006. Distributed Quasi-Monte Carlo algorithm for option pricing on HNOWs using mpC. In *Proceedings of the 39th annual simulation symposium (ANSS'06)*, Huntsville.
- Cizek, P., W. Härdle, and R. Weron. 2011. *Statistical tools for finance and insurance*. Springer.
- Cutting, D., M. Cafarella, and B. Lorica. 2016. *The next 10 years of Apache Hadoop*. O'Reilly Media.
- Dean, J., and S. Ghemawat. 2004. MapReduce: Simplified data processing on large clusters. *OSDI '04*, 137–150.
- Dowd, K. 2002. *Measuring market risk*. Wiley.

- Efron, B. 1981. Censored data and the bootstrap. *Journal of American Statistical Association* 76 (374): 312–319.
- Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1 (1): 54–75.
- Faure, H. 1982. Discrepance de Suites Associees a un Systeme de Numeration (en Dimension s). *Acta Arithmetica* 41: 337–351.
- Financial Stability Board. 2017. Artificial intelligence and machine learning in financial services-market developments and financial stability implications. *FSB Report*.
- Foster, I., and C. Kesselman. 2004. *The grid: Blueprint for a new computing infrastructure*. 2nd ed. Morgan Kaufmann.
- Foster, I., Y. Zhao, I. Raicu, and S. Lu. 2008. Cloud computing and grid computing 360-degree compared. In *Grid computing environments workshop*, 1–10.
- Galanti, S., and A. Jung. 1997. Low-discrepancy sequences: Monte Carlo simulation of option prices. *Journal of Derivatives* 5 (1): 63–83.
- Gandhi, R., and A. Jayanti. 2020. Technology factsheet: Differential privacy. *Report*, Belfer Center for Science and International Affairs, Harvard Kennedy School.
- Ghoshal, D., Canon, R.S., and L. Ramakrishnan. 2011. I/O performance of virtualized cloud environments. In *The second international workshop on data intensive computing in the clouds*, 71–80.
- Gillam, L., Ahmad, K., and G. Dear. 2005. Grid-enabling social scientists: The FINGRID infrastructure. In *Proceedings of the 1st international conference on e-social science*, Manchester, 22–24.
- Griffiths, P., and D. Remenyi. 2003. Information technology in financial services: A model for value creation. *Electronic Journal of Information Systems Evaluation* 6 (2): 107–116.
- Grimshaw, A.S., and W.A. Wulf. 1997. The legion vision of a worldwide virtual computer. *Communications of the ACM* 40 (1): 39–45.
- Halton, J.H. 1960. On the efficiency of certain Quasirandom sequences of points in evaluating multidimensional integrals. *Numerische Mathematik* 2: 84–90.
- Hauswald, R., and R. Marquez. 2003. Information technology and financial services competition. *The Review of Financial Studies* 16 (3): 921–948.
- Hochreiter, S., and J. Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9 (8): 1735–1780.
- Hull, J.C. 2003. *Options, futures and other derivatives*. 5th ed. Upper Saddle River: Prentice Hall.
- Joe, S., and F.Y. Kuo. 2003. Remark on algorithm 659: Implementing Sobol's Quasirandom sequence generator. *ACM Transactions on Mathematical Software* 29 (1): 49–57.
- Korpela, E., et al. 2001. SETI@home-massively distributed computing for SETI. *Computing in Science and Engineering* 3 (1): 78–83.
- Koschnick, T. 2008. Cluster & security: Separating application environments. In *International supercomputing conference and exhibition*, Dresden, 17–20.
- Kuipers, L., and H. Niederreiter. 1974. *Uniform distribution of sequence*. New York: Wiley.
- Lauret, J., M. Walker, S. Goasguen, and L. Hajdu. 2010. From grid to cloud, the STAR experience. In *SciDAC 2010 proceedings*.
- Leto, A., et al. 2005. EGRID project: Experience report on the implementation of a grid infrastructure for the analysis of financial data. In *Proceedings of the international conference on information technology: Coding and computing*.
- Lin, F.-P., L.-C. Chen, and C.-Y. Lin. 2020. Study of cyclic trading pattern on large network of unified invoice. *Advances in Investment Analysis and Portfolio Management* 10: 131–144.
- Litzkow, M.J., M. Livny, and M.W. Mutka. 1988. Condor-a hunter of idle workstations. In *8th international conference on distributed computing systems, proceeding*, 104–111.
- Mell, P., and T. Grance. 2011. The NIST definition of cloud computing. *Special publication* 800-145.
- Milojicic, S.D., et al. 2002. Peer-to-peer computing. In *HL Laboratories research report*.

- Niederreiter, H. 1992. Random number generation and quasi-Monte Carlo methods. In *CBMS-NSF regional conference series in Appl. Math.*
- Paskov, S.H., and J.F. Traub. 1995. Faster valuation of financial derivatives. *Journal of Portfolio Management* 22 (1): 113–120.
- Phillips, A., et al. 1984. *Effects of information technology on financial services systems*. Washington, DC: U.S. Congress, Office of Technology Assessment, OTA-CIT-202.
- Ramakrishnan, L., R.S. Canon, K. Muriki, I. Sakrejda, and N.J. Wright. 2011. Evaluating interconnect and virtualization performance for high performance computing. In *Proceedings of 2nd international workshop on performance modeling, benchmarking and simulation of high-performance computing systems*, 1–2.
- Schmidt, A. 2005. *Quantitative finance for physicists: An introduction*. Academic Press Advanced Finance Series.
- Smarr, L., and C. Catlett. 1992. Metacomputing. *Communications of the ACM Archive* 35 (6): 44–52.
- Snir, M., et al. 1996. *MPI: The complete reference*. Massachusetts Institute of Technology.
- Sobol, I.M. 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R. Computational Mathematics and Mathematical Physics* 7 (4): 86–112.
- Stokes-Rees, I., et al. 2007. Multi-cluster parallel job submission: Experiences with Monte Carlo simulations for computational finance on Grid5000. IRIS PicosouGrid Project Report.
- Strandell, E., S. Tilak, H.M. Chou, Y.T. Wang, F.-P. Lin, P. Arzberger, and T. Fountain. 2007. Data management at Kenting's underwater ecological observatory. In *Proceedings of the third international conference on intelligent sensors, sensor networks and information processing*.
- Tanaka, S. 2003. Joint development project applies grid computing technology to financial risk management. In *Nikkei Computer, special section on grid computing*.
- Tilak, S., P. Hubbard, M. Miller, and T. Fountain. 2007. The Ring Buffer Network Bus (RBNB) DataTurbine streaming data middleware for environmental observing systems. In *Proceedings of third IEEE international conference on e-science and grid computing*, 125–133.
- Weron, R. 2004. *Handbook of computational statistics*. Berlin: Springer.
- Xian, X., X. Wang, J. Ding, and R. Ghanadan. 2020. Assisted learning: A framework for multi-organization learning. In *Advances in neural information processing systems*, vol. 33. Curran Associates.
- Yang, Q., et al. 2019. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology* 10: 1–19.
- Yelick, K., Coghlan, S., Draney, B., and R.S. Canon. 2011. The Magellan report on cloud computing for science. *ASCR, DOE*.

Thomas S. Y. Ho and Sang Bin Lee

Abstract

The board of directors must now prepare risk management policies for regimes of low-interest rates and increased loss rates to fulfill its fiduciary responsibility to ensure the safety and soundness of the financial institution. Negative rates and higher unexpected loss rate regimes can no longer be ignored, and bankers should seek answers to the following crucial questions: How should the balance sheet be managed if negative rates and higher losses prevail? Are the current risk management models appropriate for managing the balance sheet?

This chapter explains how the local volatility model can provide what-if financial simulations, essential to balance sheet risk management, to address these questions. In particular, the chapter proposes financial simulation assumptions for the base case and stress scenarios. It also formulates balance sheet strategies for community banks to manage capital, liquidity, credit, and interest rate risks consistent with the bank's risk culture while maintaining profitability. Furthermore, it demonstrates the validity of the local volatility model to navigate the impacts of COVID 19, allowing bankers to model balance sheet elements affected by the capital market's negative rate regimes. Finally, it highlights the limitation of interest rate models widely used today in dealing with the applications described above.

The authors would like to thank Lindsay Elizondo, Gnanesh Coomaraswamy, and Blessing Mudavanhu for their comments and suggestions.

T. S. Y. Ho (✉)
Thomas Ho Company, Ltd., New York, NY, USA
e-mail: Tom.Ho@ThomasHo.com

S. B. Lee
Hanyang University, Seoul, South Korea

Keywords

Rate distribution · Forward rates · Distribution skewness · Arbitrage-free models · Interest rate models · Normal modes · Lognormal models · Interest rate forecast · Calibration · Expectation hypothesis · Fisher equation · Local volatility model · Term structure of volatilities

81.1 Introduction: Modeling Interest Rate Movements in a Low-Rate Regime

A negative rate regime in the USA has become probable because of COVID 19. On March 24, 2020, the 3-month treasury rates fell to 1 basis point, with the potential to fall further. Additionally, some secured overnight financing rates (SOFR) were negative between June 29 and July 15, 2020. Furthermore, most developed economies have negative government rates.¹ The US positive rate regime is, in fact, an exception.

COVID19 has also resulted in higher expected and unexpected credit losses. The unemployment rate was rising to 15%, falling to 12% by June 2020, even before the impact of resurgence. Such an economic downturn can lead to a deep recession and deflation, leading to negative projected asset returns.

Community bankers should simulate what-if pro forma financial statements in light of COVID 19 and reevaluate the validity of the interest rate models to perform such simulations. Financial Institution Board members have the fiduciary responsibility to raise questions such as:

- What are the balance sheet strategies to manage negative rate and higher loss rate scenarios?
- Is the current risk management model appropriate for managing the balance sheet?

This chapter addresses these questions in two parts:

Part 1 describes the impact of COVID 19 on community bank performance. First, I propose the specifications of financial simulations, including what-if analyses to identify risk exposures. Second, based on the simulations, I suggest balance sheet strategies using the latest research. Finally, I show that the efficacies of these pro forma financial analyses depend on the appropriateness of the interest rate model used.

Part 2 describes the interest rate model that remains robust in the COVID 19 environment. I also identify the importance of the interest rate model to implement the proposed what-if scenarios and strategies to manage credit, interest rate, liquidity, and capital risk in a holistic modeling framework.

¹As of July 15, 2020, the 2-year government interest rates (basis points) of the following countries are France-59, Germany-66, Italy-6, Japan-0, 14, Spain -40, and the UK-10 (Sources: WSJ).

81.2 Part 1: Model Assumptions for Financial Simulations and Balance Sheet Strategies

81.2.1 COIVD 19: Challenges and Strategic Planning

The capital market does not expect the near-zero interest rate policy in the USA, nor the global negative interest rate policy (NIRP) will soon end. These policies' negative rates will "flow through to the markets."² However, the Federal Reserve Bank (San Francisco) chapter experts banks to "intermediate the transmission of the Fed policy" on the COIVD 19 shock, and the impact will evolve.

COIVD 19 will affect community banks with less diversified funding and products differently than wholesale banks. Therefore, it is more pertinent for community banks to "create and analyze bespoke shat-if scenarios" for the COIVD 19 regime, as PWC recommends (2020). In particular, competition to lower the funding rate to maintain the margin on the liability side as the lending rates are compressed.

81.2.2 Model Assumptions

THC's research suggests that the assumptions for the base case scenario should keep the current interest rates unchanged for two years, with increased projected loss rates. The level of increase will depend on the loan portfolio's credit exposure and borrower's risk profiles. The nonmaturity deposit account rates maintain a reasonable level to compete with wholesale banks and fintech companies.

Community banks should also simulate stress scenarios. Since the local volatility model estimates the probability of the 1-year rate becoming negative in five years is over 35%, the model suggests that a 100 basis point downward rate shock is reasonable as a stress test. Even if the bank resists offering negative deposit rates, the capital market rates will affect the pricing of products such as ARMs, where indices are often market based, such as the LIBOR or its replacement. However, to manage profitability, bank rates may not fall in step with the market rates, so margins may not be affected as significantly. Other developed economies have confirmed such relationships between capital market rates and loan rates. The stress scenario will also increase credit losses. The board and senior management should revise the performance dashboard based on both the base case scenario and the stressed scenario.

81.2.3 Balance Sheet Strategies

While these scenarios will adversely impact community banks' performance, other economies have shown that the results can be manageable, when senior management

²PIMCO (2019).

takes appropriate actions. McKinsey (2020) recommends banks should be “taking a more holistic approach to using (balance sheet management) tools.” Their survey suggests that banks should increase the balance sheet by originating profitable products. The emphasis is on identifying the most profitable products. When capital constrains the growth of the balance sheet, the fee-based business should increase.

Loan profitability should be holistically determined. The revenue is net of not only the funding cost, but also the borrowers’ option cost, credit risk charges, and noninterest expenses to support the products. The activity-based cost analysis can estimate the cost assigned to each product. Increasing capital via retained earnings is essential in this environment, as risk managers should note that the full impact of COIVD 19 is evolving and will not be felt immediately (Arseneau 2017).

In essence, based on a survey (see Alberts 2020), these bank strategies are acting as “transmission channels” for the Federal Reserve Bank Interest Rate Policy that seeks to increase the availability of credits to the real economy, by increasing lending and lowering the credit risk premium. The proposed strategies are consistent with the mechanics of banking, acting as the “transmission channel.”

81.2.4 The Validity of Interest Rate Models

The validity of financial simulations in the COIVD 19 environment, of course, will depend on the robustness of the interest rate model in the low-interest rate regime, and therefore, I must first address the validity of the interest rate model.

Interest rate models are central to valuing the embedded fixed-income options, which are prevalent in balance sheets, and are essential to enterprise risk management and market making. They are also the bedrock to building credit and liquidity models, as well as income/profitability simulations.³ Financial economists cannot overstate the importance of interest rate models to our economic system. However, many asset-liability models were developed in the 1990s and are not appropriate for today’s environment.

The current low-interest-rate regime challenges the robustness of many interest rate models.⁴ These models allow interest rates to fall to a prespecified level or to rise unreasonably high. In sum, these models fail to provide reasonable interest rate distributions, which are determined by capital market prices.

By way of contrast, the chapter describes the local volatility model, which allows the capital market prices to determine the interest rate distribution. That is, the local volatility model is “arbitrage-free,” analogous to level II fair value accounting principle in relative pricing. Hence, the local volatility model responds to changing market conditions and perceptions of interest rate uncertainties. To illustrate, the chart below depicts a rate probability distribution of the local volatility model. The

³See Ho and Coomaraswamy (2019).

⁴See Ho and Mudavanhu (2019).

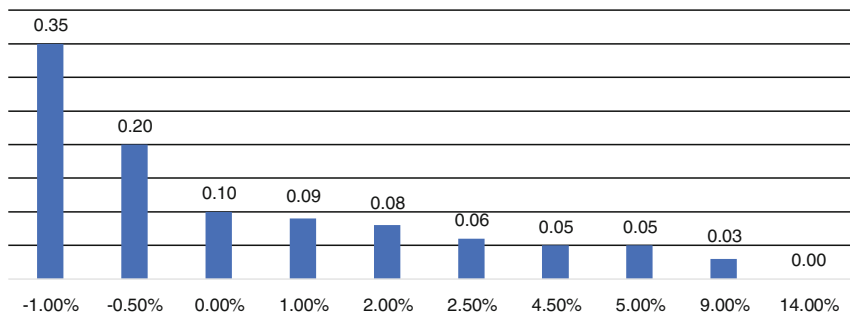


Fig. 1 Probability distribution of 1 month rates at a 5 year horizon (March 6, 2020)

results show that the probability of the rate falling between -0.5% and -1% is 35% the probability for negative rates, in general, is 55% (Fig. 1).

The probability distributions estimated by the local volatility model enables senior management to measure the impact of COIVD 19 on balance sheet risk and the efficacy of balance sheet strategies.

81.3 Part 2: Interest Rate Model and Efficacy of Financial Simulations

In this section, we summarize the chapter that presented the local volatility model and its applications in 2019. We will provide an intuitive explanation of the model and the implied probability distributions of interest rates based on a sample of months from March 2019 to February 2020. From the empirical results, we will discuss several balance sheet strategies. we will further extend the discussion to financial-economic issues. Finally, we will introduce new approaches to balance sheet total return, lending, and funding strategies.

81.3.1 Local Volatility Model

Interest rate models should accept negative interest rates implied by swaption prices and asymmetric distribution of the frequency distribution of interest rates. The local volatility models such as Ulate (2019), Cox et al. (1981), and Ho and Lee (1986) follow the Ho-Lee formulation (see Appendix A) by not assuming a specific interest rate probability distribution, such as the normal or lognormal distributions. Instead, the local volatility model assumes the volatility is local, depending on the time horizon and the projected rate levels, local in the projected time, and rate level.

The model calibration uses both at-the-money and out-of-the-money swaptions to calibrate the interest rate movements and hence the rate distributions. The local volatility model uses the 80 at-the-money swaptions with terms 1, 2, 3, 4, 5, 7, and 10 years and tenor 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, and 91 out-of-the-money

swaptions with rates -100 , -50 , 50 , and 100 bp, for the same term and tenor. The local volatility model can then present the underlying rate distributions. In this sense, the local volatility model can then present the underlying rate distributions. In this sense, the local volatility model is arbitrage-free, as it is consistent with the capital market pricing of the yield curve and observed swaption prices.

Figure 2 below depicts the 1-month treasury rate distribution over a 5-year horizon as of March 6, 2020, and March 13, 2020, when the coronavirus pandemic greatly affected the treasury and swaption markets.

The results show that there is a significant implied probability of negative interest rates prevailing in 5 years. Indeed, the results indicate there is a 35% and 20% chance the 1-month rate will reach less than -0.5% and fall between -0.5% and 0% , respectively.

Three risk drivers are affecting the rate probability distribution: (1) the forward rate (2) market-perceived interest rate uncertainty, called the volatility curve, and (3) the out-of-the-money swaptions. These risk drivers determine the median, variance, and skewness of the distribution.

The following section describes the historical trends of the rate distributions as affected by these three risk drivers.

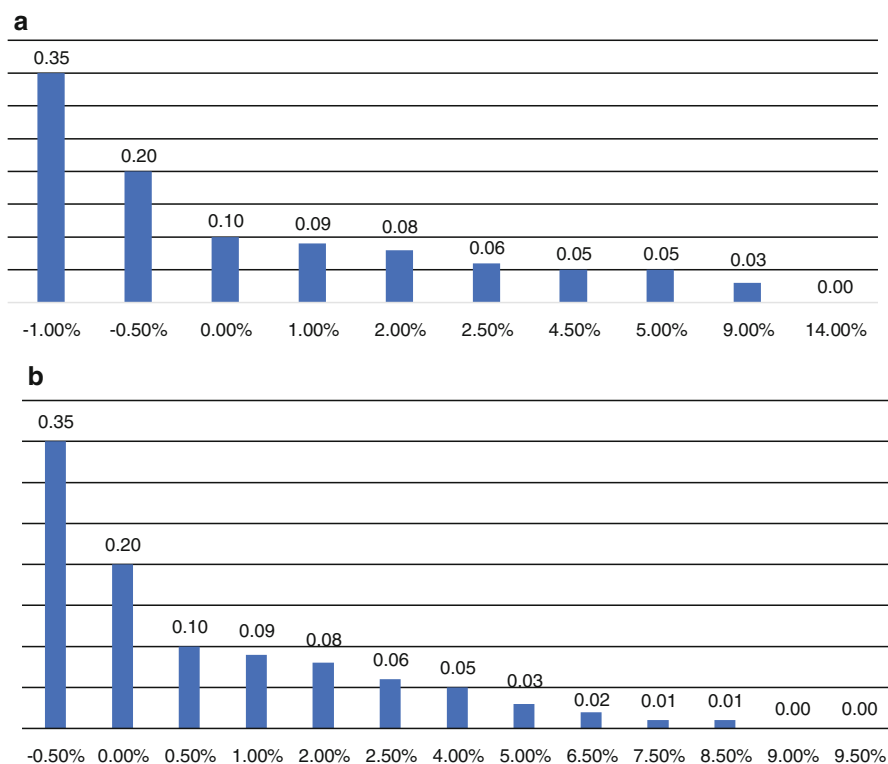


Fig. 2 Probability distribution of 1 month rate at a 5-year horizon. (a) March, 6 2020, (b) March, 13 2020

81.3.2 Rate Distribution Trend Analysis

81.3.2.1 Market Conditions

The shape of the yield curve determines the forward rates. The historical trends of the key rates, shown below, can partly explain the median rate of the historical rate distributions.

The 12-month daily treasury key rates depict the dynamic nature of the rate level and the shape of the yield curve. Via the swaption market, the perceived uncertainty and skewness of the rate distribution can provide useful information on a forward-looking basis.

81.3.2.2 Volatility Curve

The local volatility model uses swaptions to determine the uncertainty of rates along the yield curve. Figure 4 shows the market's perceived rate uncertainty. As expected, the market volatility was high during the post-2008 financial crisis, and again when the Brexit event occurred in 2011. The volatility again rose in the period of the trade war and more currently during the coronavirus pandemic.

81.3.2.3 Negative Rates

Given the fall in interest rates, increased volatility, and the change in the skewness of the distribution, the implied probabilities of negative interest rates change over the sample period. Below is the monthly estimate of the rate distributions over 12 months starting at the end of March 2019. The results show that August 2019 and March 2020 have significant implied negative rates.

The most substantial one-day losses of 2019 occurred in August after we saw significant market uncertainty as China suggested using their currency to strategize the trade war. Still, the risk subsided after the trade discussion resumed in the following months.

By December, the coronavirus outbreak in China became widespread, resulting in a rise in global market performance uncertainty. The risk heightened when the spread of the virus became a pandemic in March. Comparing the vol curve and rate trends, the significant rise in the probability of a negative rate is mainly due to the drop in the rate level, not as much as the increase in the normal rate risk.

81.3.3 Model Robustness: Historical Rate Distributions for Sample Dates Between March 2019 and March 2020

On March 29, 2019, the treasury curve was relatively flat, suggesting the projected forward rates were not high, and the curve level was high relative to subsequent months. The implied volatility was also low before the increase in volatility that resulted from the trade war uncertainty.

81.3.3.1 Empirical Evidence

The results reported in Figs. 3, 4, 5, and 6 show that the implied distribution is consistent with these observations. The distribution model is around 1%, with the rates not falling below zero. As a result, the probability distribution is positively skewed, with a median of 2%.

The following Figs. 6, 7, 8, and 9 further support these observations based on additional sample data.

As explained above, because of the escalation of the trade war, August 2019 experienced significant market uncertainty. Figure 3 shows that interest rates rose

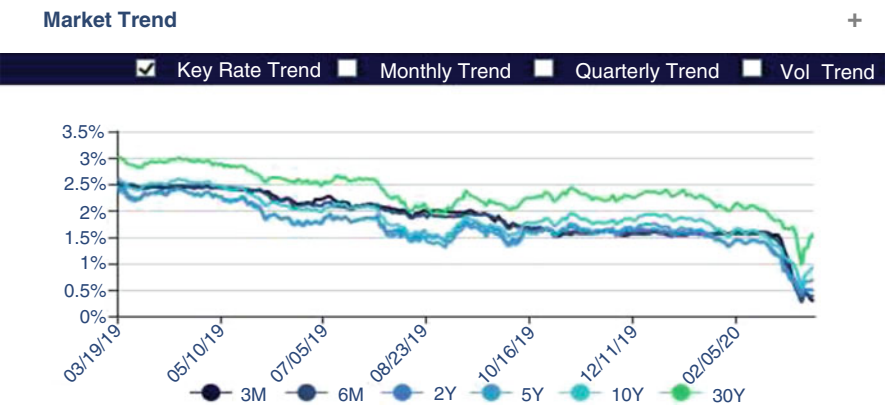


Fig. 3 The historical trends of treasury key rates



Fig. 4 Historical volatility curves on interest rate risks. The results show that the short-term uncertainty is more pronounced in this coronavirus pandemic period, as reflected in the rise of the short-term rate volatilities. Note that while recent volatilities are lower than those in 2010, the reason is that the interest rates have fallen substantially since 2010. If the interest rate risk is measured by percentage change (lognormal model), then the current volatility is over 1500%.

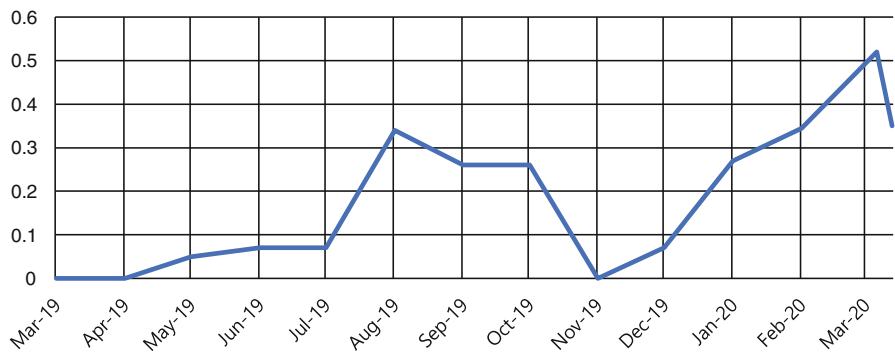


Fig. 5 Historical implied probabilities of negative interest rates. While most periods do not exhibit a positive probability of negative interest rates, the swaption markets with both at-the-money and out-of-the-money swaptions identify the likelihood of negative rates in stress periods

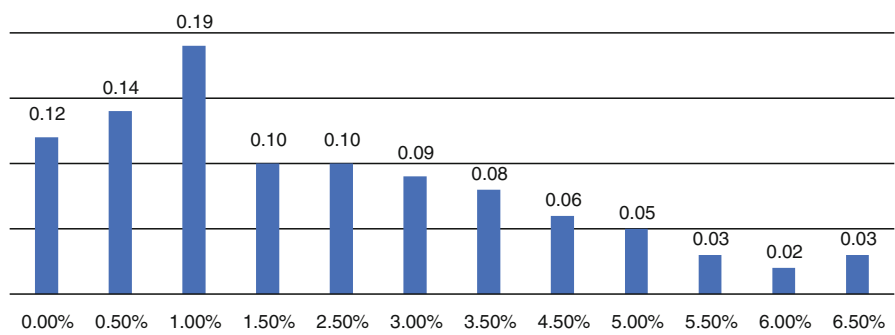


Fig. 6 Probability distribution of 1 month rate at a 5-year horizon. (March, 29 2019)

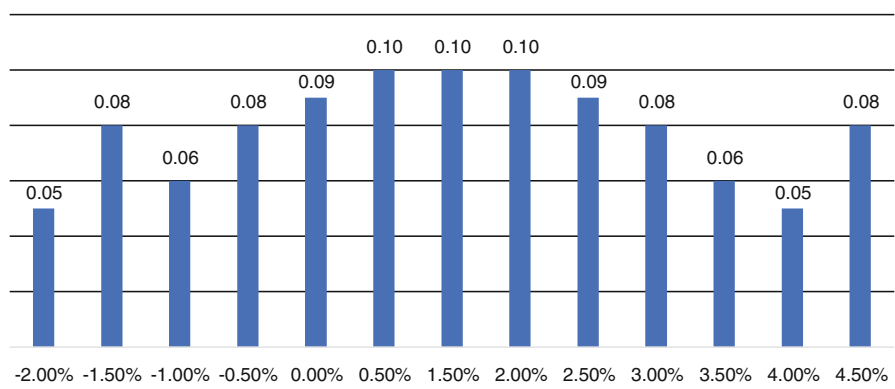


Fig. 7 Probability distribution of 1 month rate at a 5-year horizon. (September 30, 2019)

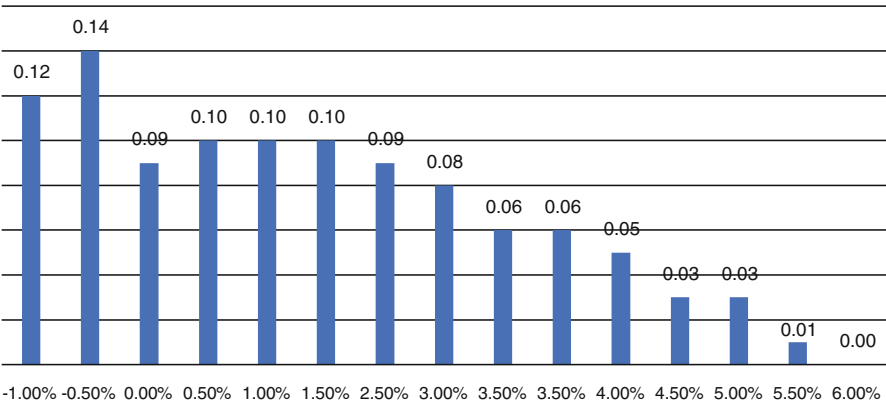


Fig. 8 Probability distribution of 1 month rate at a 5-year horizon. (October 31, 2019)

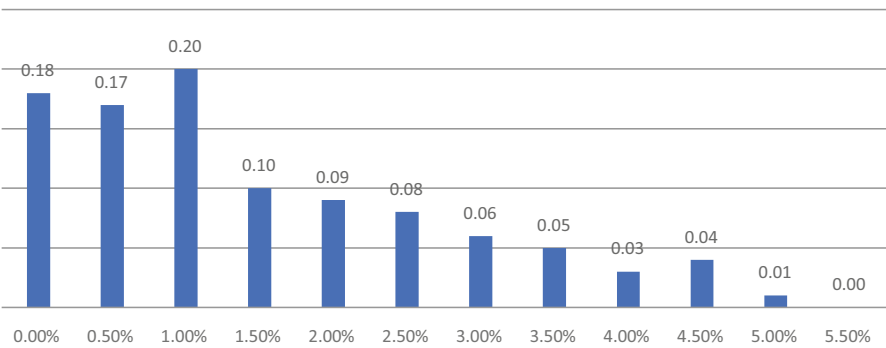


Fig. 9 Probability distribution of 1 month rate at a 5-year horizon. (November 29, 2019)

significantly, and Fig. 4 shows that the implied volatility increase in the same period. Figure 6 shows that the median of the distribution shifted higher, and the rates fell below zero.

In late October, the yield curve is no longer inverted, and volatility fell. Figure 9 shows that the rate of distribution has a lower probability of negative rates, and the skewness is more pronounced than in the previous month.

By late November, the volatility fell further, and the yield curve remained flat around 1.75%. Figure 8 shows that the rate distribution does not have negative rates, and the skewness remained positive.

The results show that:

- The dynamic nature of the distribution indicates that the median, variance, and skewness of the distributions continually adjust to the market conditions as the yield curve, implied volatilities of the at-the-money swaptions, and the skewness affected by the out-of-the-money swaption pricing change over time.

- A lognormal distribution cannot model the distribution because lognormal distributions do not accept negative interest rates. When the interest rate is only 0.2%, the rate distribution is constrained to a narrow band on variations. Therefore, the projected simulations cannot be consistent with the swaption prices.
- The normal distribution cannot be consistent with the current low-interest-rate scenario as a result of the market “Risk Arbitrage” practice. If rates become significantly negative, investors have alternative investments without paying a significant price premium that results in high negative interest rates. For these reasons, the historical estimated distributions tend to be positively skewed, and there is a limit to the level of negative rates.

In summary, the results show that the rate distribution is dynamic, continually adjusting to the market perception of interest rate uncertainties and the skewness of the distribution. Many operating interest rate models estimate the skewness and floor rate separated from the model calibration estimation process.⁵ Additionally, some interest rate models are calibrated to at-the-money options only, excluding out-of-the-money options. The next section will discuss the limitations in fixed-income pricing when the extent of projected negative rates and rate distribution skewness is not dynamically estimated.

81.4 Implications of Skewness: Mean and Median of Rate Distribution

The previous sections show the importance of evaluating the rate distributions and the dynamic nature of the rate distribution, neither confined to a normal or lognormal distribution. The distributions should be defined by the mean, variance, and skewness.

81.4.1 Forward Rates and Expected Rates

Rate distribution skewness concludes that the forward rates, as determined by the yield curve, are not the market’s expected rates. The forward rates and expected rates are different, contrary to the expectation hypothesis. The rate distribution can explain the poor performance of the forward rates as a rate forecasting tool and offer an alternative to measuring the market expectation of future rates.

A market scenario can best explain the importance of using rate distribution. On March 20, 2020, the yield curve was upward sloping. The 1-month and 10-year rates were 0.04% and 0.92%, respectively. However, the capital market did not have to believe the 1-month rate would rise from 0.04% to 0.51%, or the 10-year rate from 0.92% to 1.06%, in one year. Instead, the capital market traded the caps to reflect the

⁵Calibration is explained in the White Paper 2019 “Embedded Option Pricing.”

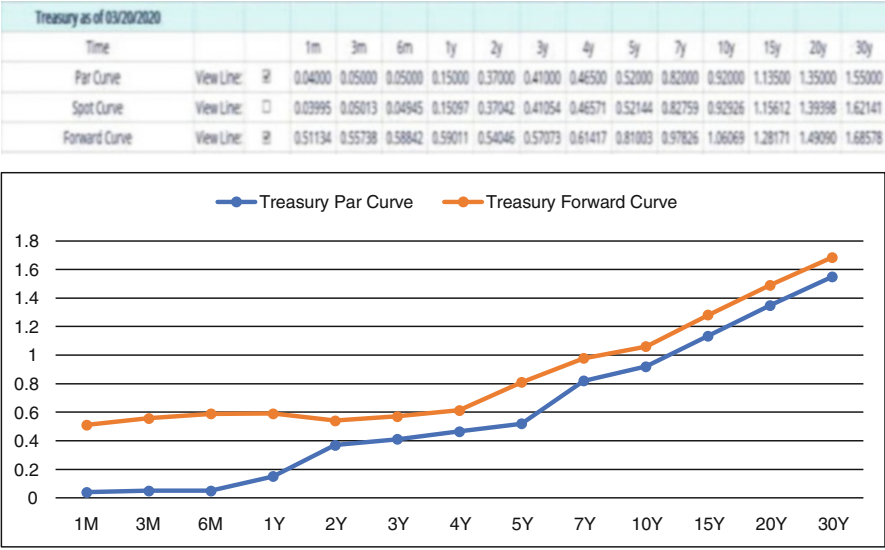


Fig. 10 The difference between the par curve and the forward curve

market rate expectation, creating a negative skewness in the rate distribution. At the same time, the Federal Reserve Bank affected a steep yield curve for a broader economic purpose.

Figure 10 presents the difference between the par curve and the forward curve and the relevance of the market interest rate forecast (MIRF).

The average rate distribution, as depicted in Fig. 2, represents the market expectation, which is not the same as the forward rates, which intuitively is the median of the distribution.

81.4.2 The Skewness in Loan Pricing

The depiction of the rate distribution has many applications in the capital market. The model provides an estimation of the value of caps and floors of adjustable rate instruments as the rate distribution reports the probabilities of becoming in the money for the caps and floors. Consider the following example:

A 30-year adjustable-rate mortgage (ARM) may have an interest rate structure of 2/2/5, indexed to a 1-year treasury rate with a margin of 2.2%. If the loan has an initial rate of 3.25%, then the cap at the reset would be useful when the 1-year treasury rises to 5.25 ($3.25\% + 2\%$), or the 1-year treasury rate increases from today's rate of 0.8% to 3.25% in 5 years. The rate distribution provides the probability of the interest to reach the cap at 5.25%. If that scenario is not likely to prevail, the lending officer could lower the initial rate so that the ARM is competitive to the rate of the 30-year fixed-rate mortgage.

81.4.3 Stress Test the Balance Sheet and Limitations of Current Interest Rate Models

The distributions also provide valuable information for stress testing the balance sheet or a portfolio of balance sheet instruments. The results provide some measure of the likelihood of a stress scenario that may occur. The “likelihood” is estimated from capital market prices and is, therefore, objective and transparent. The local volatility model can simulate the balance sheet under the downshift of the yield curve for stress testing purposes.

Researchers developed many interest rate models in the 1980s when interest rates were much higher than those of today.

- The lognormal model cannot accept negative interest rates, and therefore, will fail for the stress test.
- The normal model accepts negative interest rates, but the projected model interest rates could be significantly negative. However, rates cannot be significantly negative when employing the risk arbitrage activities of “putting money under the carpet.” Any assumption of flooring the rate at a particular level would violate the arbitrage-free model assumption. That is, such a model cannot be consistent with the basic discount cash flow model when the embedded option value is negligible and hence would misprice the balance sheet items.
- Another commonly used model, the displaced diffusion model, specifies the minimum interest rate base on a lognormal model. Empirical evidence has shown that the minimum rate is dynamic, and the floor rate has to be estimated by the calibration.

81.4.4 Balance Sheet Enhance Return Strategies

The local volatility model has introduced the concept of a rate probability distribution. From the rate probability distribution, interest rate strategists can manage with mean, variance, and skewness of the probability distributions. To date, controlling the rate level risk and interest rate volatility is widely used in practice. In the low-interest-rate regime, balance sheet strategies should also focus on skewness.

Specifically, the pricing of caps and floors, and the use of out-of-the-money options are essential for hedging, buying, or selling options. There are many out-of-the-money options sold to customers on the balance sheet. The value of skewness of the interest rate probability distribution can be perceived differently from those priced in the capital market. The variance in the price offers profit opportunities.

81.5 Conclusions

Financial economists have categorically rejected the possibility of negative interest rates for over 35 years, resulting in much unwarranted research and the development of interest rate models. Today, the low-interest-rate regime has “broke(n) the Black

Scholes model, Pillar of Modern Finance.” Economists should critically evaluate the current operational interest rate models.

This chapter explains the local volatility model and provides empirical evidence of the validity and applications of the model to manage the impact of COIVD 19 on community banks. In particular:

- **Distribution skewness:** The interest rate model should use both capital market pricing of at-the-money and out-of-the-money options to define the rate distribution, the projected interest rate minimum, and the distribution skewness. The skewness affects the pricing of caps and floors of an adjustable-rate instrument. Furthermore, interest rate models without the calibration of the interest rate distribution do not provide a valuation model consistent with capital market derivative markets.
- **The expectation hypothesis:** The interest rate model’s rate distribution determines the market expected interest rate level, which is a more appropriate rate forecast than that suggested by the forward curve, as the expectation hypothesis suggests.
- **The Fisher equation.** The concept of the nominal interest rate is the sum of the economy’s real rate of return, plus the inflation rate (the Fisher equation) is a positive theory. By way of contrast, the arbitrage-free model is a normative theory. The arbitrage-free model can lead to timely actionable decisions and not relying on the concept of general market equilibrium.

This chapter explains the local volatility model and provides essential balance sheet applications in the COIVD 19 environment. In particular, the chapter

- Proposes financial simulation assumptions for the base case and stress scenarios.
 - Projected interest rates over 2 years.
 - Increasing loss rates over the simulation horizon
- Formulates balance sheet strategies for community banks to manage capital, liquidity, credit, and interest rate risks consistent with the bank’s risk culture while maintaining profitability.
- Demonstrates the validity of the volatility model to navigate the impacts of COIVD 19, allowing bankers to model balance sheet elements affected by the capital market’s negative interest rate regimes.

Highlights the limitation of interest rate models widely used today in dealing with the applications described above.

Appendix A: Interest Rate Model: A Review

Ho-Lee Model

Ho and Lee 1986 introduced the arbitrage-free interest rate model in 1984. “Arbitrage-free” means that the evolution of the yield curve movements must be consistent with the valuation of capital market benchmark derivatives and the market discount curve.

Swaptions and the swap curves are examples of derivatives and market discount curves, respectively. I called the process of implying the underlying yield curve movements from the derivative prices “Calibration.” (Ho (2019a)) Since the derivative prices determine the yield curve movements by calibration, by definition, there is no arbitrage opportunity in the capital market to exploit the pricing of derivatives. The Ho-Lee result shows that the calibrated probability distributions of interest rates can take on many forms depending on the time horizon and the selected rate to be considered.

Normal Model Versus Lognormal Model

When interest rates were high, around 6%, in the 1980s, interest rates tended to rise and fall by the same amount. The normal model (Ho and Lee (1986)) hypothesizes that rates are equally likely to rise and fall, which is approximately consistent with the swaptions’ prices. That is the normal model measures the rate change by the basis point change, a basis point shift. However, the normal model allows for negative interest rates, an assumption that financial economists categorically rejected. In response, the financial modelers replaced the normal model with the lognormal model.

The lognormal model (Black et al. 1990) assumes interest rates take random walks, analogous to that of the stocks, which always remain positive. The model measures change by the percentage change. When rates are low, the percentage of change can be high. The lognormal model projects rates rising rapidly, creating a positive skewness of the rate distribution, an undesirable attribute.

Nearly 35 years of research sought to resolve these inconsistencies with observed market prices. The figures below illustrate the complexity of interest rate simulations by comparing the outputs of a lognormal model and a normal model. The charts depict the monthly 1-month swap rates simulated over 30 years. The BGM model,⁶ an operational interest rate model, generates 257 interest rate paths under a lognormal (on the left) and normal model (on the right). These interest rate paths are consistent with market pricing of the swaps and a portfolio of swaptions, satisfying the arbitrage-free model assumption.

The simulation shows that the lognormal model does not allow negative interest rates, resulting in a positively skewed rate distribution. By way of contrast, the normal model allows for significantly negative interest rates while maintaining the rate distribution to be symmetric (Figs. 11 and 12).

Appendix B: Local Volatility Model Description

Financial economists have made tremendous progress in interest rate modeling in the past 35 years. Models such as Ho and Lee (1986), Black et al. (1990), Cox et al. (1981), Hull and White (1990), Black and Karasinski (1991), Brace et al. (1997), and

⁶Reference: Technical Notes explains the BGM model.

Figure 11 Underscores the limitation of these interest rate models. Today, interest rate frequency distribution is neither lognormal nor normal. The former does not accept negative rates. The latter cannot take out-of-the-money option prices to determine the asymmetric cap and floor premiums and often leads to significant negative interest rates.

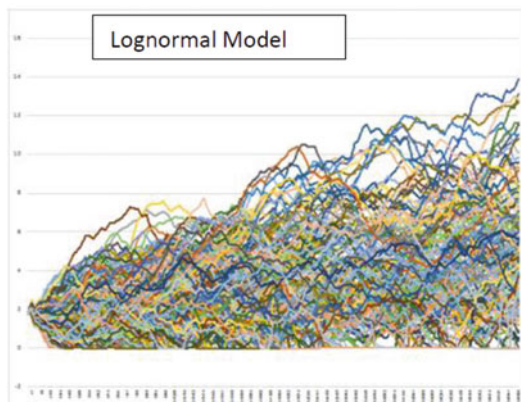
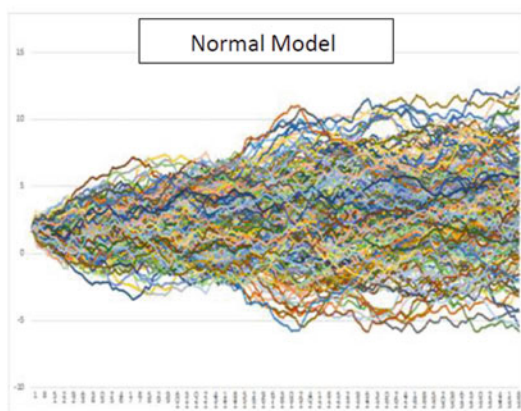


Figure 12 Presents interest rate paths over 30 years. The results depict the limitations of the lognormal and normal interest rate models. The lognormal model is positively skewed with the mean 1-month rate that exceeds the median rate at a time horizon. The normal model has no skewness with the average rate and the median rate the same.



Longstaff et al. (2002) continually enhance interest rate modeling. These models have also introduced many new concepts such as martingale, lattice, recombining, delta hedge, risk-neutral measure q , physical measure p , kernel, vol surface, lognormal and normal models, OAS and Greeks, string theory, stratified sampling, rational option exercise rules, and CEV skew model. The study of interest rate modeling has even become a core course in the mathematics department.

For the following equations, we use the notations:

r = a short – term rate

$\emptyset$ = term structure of volatilities.

$\sigma(t)$ = term structure of volatilities

dz = wiener process; normal distribution over a short time

The basic models from which extend many other models for credit risks yield curve movements, computational efficiencies:

$$dr = \emptyset(t)dt + \sigma(t)dz$$

The term structure of volatilities is independent of the rate level in the normal model. The Heath et al. (1992) models determine the term structure of volatilities using the forward rates.

$$dr = \emptyset(t)rdt + \sigma(t)rdz$$

The term structure of volatilities is proportional to rates in the lognormal model. Black et al. (1990) avoids negative rates.

$$dr = \emptyset(t)(1-r)dt + \sigma(t)r^{0.5}dz$$

The Cox et al. (1981) model shows where the rates mean revert to some long-term rate and rate change depends on the term structure of volatilities and positively related to the rate level

$$dr = \emptyset(t, r)dt + \sigma(t)r^{\alpha}dz$$

The elasticity (α) of the constant elasticity variance model is between 1 and 0.1. BGM (LIBOR market model) is a CEV model with an efficient algorithm that fits precisely to market pricing of swaptions. These models do not accept negative rates. To allow for negative rates, BGM and Hull-White models allow for a modeler to set the floor rate, which can be negative.

$$dr = \emptyset(t)dt + \sigma(t)\eta(r)dz$$

The local volatility model allows for the out-of-the-money options to specify the rate distribution enabling the rate distribution $\eta(r)$ to fit the out-of-the-money option prices (Ho and Mudavanhu 2019).

The established interest rate models do not allow for the swaption prices to specify the rate distribution. Instead, these models impose the swaptions be priced based on a form of lognormal models or a normal model with a constraint of the minimum rate. The local volatility model seeks to overcome these problems in a low-interest-rate regime by explicitly determine the rate distribution $\eta(r)$ from the OTM option prices.

References

- Alberts, M. 2020. *Negative rates and insurance industry survey of risk-management capabilities and practice*. Casualty Actuarial Society, Society of Actuaries: Canadian Institute of Actuaries. Available at <https://www.soa.org/globalassets/assets/files/resources/research-report/2020/negative-interest-rates.pdf>.
- Andreas, B., Olivier, P., Thomas, P., and Sebastian, S. 2020. *How banks can protect margin from negative interest rates*. McKinsey.

- Arseneau, D.M. 2017. *How would US banks fare in a negative interest rate environment?* Finance and economics discussion series 2017–030. Washington: Board of Governors of the Federal Reserve System. Available at <https://doi.org/10.17016/FEDS.2017.030>.
- Baird, S., A. Schwartz, and T. Yue. 2020. *Managing through uncertainty: Recommendations to bank treasurer*. White Paper, New York: PWC. Available at <https://www.pwc.com/us/en/library/covid-19/recommendations-for-bank-treasurers.html>.
- Black, F., E. Derman, and W. Toy. 1990. A one-factor model of interest rates and its application to Treasury bond options. *Financial Analysts Journal* 46: 33–39.
- Black, F., and P. Karasinski. 1991. Bond and option pricing when short rates are lognormal. *Financial Analysts Journal* 47: 52–59.
- Brace, A., D. Gatarek, and M. Musiela. 1997. The market model of interest rate dynamics. *Mathematical Finance* 7: 127–155.
- Cox, J., J. Ingersoll, and S. Ross. 1981. A reexamination of the traditional hypothesis about the term structure of interest rates. *Journal of Finance* 36: 769–799.
- Heath, D., R. Jarrow, and A. Morton. 1992. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica* 60: 77–105.
- Ho, T.S.Y. 2019a. Rate surface: Managing option risk. In *White paper*. New York: Thomas Ho Company.
- . 2019b. *Market interest rate forecast*. White Paper, New York: Thomas Ho Company.
- Ho, T.S.Y., and G. Coomaraswamy. 2019. *pOAS: Profitability measure for loan acquisition and loan portfolio management*. White Paper, New York: Thomas Ho Company.
- Ho, T.S.Y., and S.B. Lee. 1986. Term structure movements and pricing interest rate contingent claims. *The Journal of Finance* 41: 1011–1029.
- Ho, T.S.Y., and B. Mudavanhu. 2019. *Arbitrage-free interest rate model in negative interest rate regime*. White Paper, New York: Thomas Ho Company.
- Ho, T.S.Y., and B. Mudavanu. 2007. Interest rate models' implied volatility function stochastic movements. *Journal of Investment Management* 5: 1–22.
- Hull, J., and A. White. 1990. Pricing interest rate derivatives securities. *Journal of Financial Studies* 3: 573–592.
- Longstaff, F., P. Santa-Clara, and E.S. Schwartz. 2002. The relative valuation of caps and swaptions: Theory and empirical evidence. *Journal of Finance* 56: 2067–2109.
- PIMCO. (2019). "Understanding investing in a negative interest rate." World education series. Available at <https://global.pimco.com/en-gbl/resources/education/investing-in-a-negative-interest-rate-world>.
- Ulate, M. 2019. *Going negative at the zero lower bound: The effects of negative nominal interest rates*. Federal Reserve Bank of San Francisco: Working Paper. Available at <https://doi.org/10.24148/wp2019-21>.



Applications of Logistic Regression and Hazard Method in Accounting and Finance Research

82

Feng Gao

Abstract

This chapter introduces commonly used models to study whether or when an event occurs in accounting and finance research. I begin with parametric models based on the different survival time assumptions and then move on to a semi-parametric regression model, i.e., the Cox proportional hazards model. Finally, I introduce the use of logistic regression in discrete survival time. After reviewing prior literature using these methods, I use an empirical setting to illustrate the applications of these methods in SAS and compare the inferences.

Keywords

Logistic regression · Hazard regression · Survival analysis

JEL Classification

C24 · C34 · C41

Abbreviations

PDF propensity density function

82.1 Introduction

Hazard models are widely used statistical methods that are used to analyze data in clinical trials. Researchers are interested in the time until the occurrence of an event, such as being cured of diseases or death – these events are one-time events. These

F. Gao (✉)

Department of Accounting and Information Systems, Rutgers University, New Brunswick, NJ, USA

e-mail: fgao@business.rutgers.edu

models are also used to estimate events that happen multiple times, like the relapse of diseases. These models have been applied to research areas other than medical research, such as economics, finance, and accounting.

In accounting and finance research, there is often a need to estimate the risk of certain notable events, including the failure of firms (i.e., Shumway 2001; Iwasaki 2014; Bhattacharjee et al. 2009; Gerakos et al. 2013; Fjesme et al. 2021), small business owners (e.g., Agarwal et al. 2005), hedge funds (e.g., Lee and Kim 2014), or banks (e.g., Audrino, Kostrov, and Ortega 2019). Researchers are also interested in the factors leading to personnel turnover (e.g., Jenter and Kanaan 2015; Maturana and Nickerson 2020).

Researchers also predict the likelihood of bond default (e.g., Moeller and Molina 2003), the beginning of foreclosures (e.g., Agarwal et al. 2018), the termination of an initial public offering (e.g., Chen, Guo, and Lin 2010), the early payoff or default of personal loans or credit card accounts (e.g., Banasik et al. 1999; Bellotti and Crook 2009), or the likelihood of regulatory actions (e.g., Leverty and Grace 2018). These studies share one common feature – the variable of interest is whether events would occur and when they occur. The statistical methods used include various hazard regression models and logistic regression. In this section, I briefly introduce the theory and application of commonly used models. In later sections, I first review the empirical literature in more detail and then use the data on class action lawsuits and whistleblowing to illustrate the application of different hazard models.

82.1.1 Hazard Models

82.1.1.1 Hazard Models: Some Basic Concepts

A key concept in hazard models is the survival time until an event occurs (T). When the time to event T is a continuous random variable, its cumulative distribution function is¹

$$F(t) = \Pr(T \leq t) = \int_0^t f(u) du.$$

The probability density function (pdf) of T is

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}.$$

The survivor function is

$$S(t) = \Pr(T > t) = 1 - F(t).$$

¹I use the notation from Gao and He (2020) throughout the paper.

By definition, $S(0) = 1$, $S(\infty) = 0$. This function measures the probability of survival at time t if a subject is alive at the origin of time ($t = 0$), and it is decreasing in time.

The hazard function, i.e., an instantaneous risk that an event will happen at time t , is defined as

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

The hazard function is not a probability; instead, it is an instantaneous risk that an event will happen at time t . The hazard function is often interpreted as the expected number of events in a one-unit interval of time, so it is also called the hazard rate. The reciprocal of the hazard rate is sometimes called the Mills ratio (Mills 1926), which gives the expected length of time until an event happens. The relationship between these functions can be expressed in multiple formats.

$$\lambda(t) = \frac{f(t)}{S(t)}$$

or

$$\lambda(t) = -\frac{d}{dt} \log S(t),$$

or

$$S(t) = \exp \left[-\int_0^t \lambda(u) du \right].$$

The cumulative hazard function is

$$\Lambda(t) = \int_0^t \lambda(u) du.$$

The relationship between the survivor function and the cumulative hazard function can be rewritten as

$$S(t) = \exp [-\Lambda(t)],$$

and the relationship between the probability density function, hazard function, and the cumulative hazard function is

$$f(t) = \lambda(t) \exp [-\Lambda(t)].$$

The most commonly used methods to estimate the survivor function are Kaplan-Meier estimator and life-table estimator. The Kaplan-Meier estimator is also known as the product-limit estimator. It is a nonparametric statistic that was independently

developed by Kaplan and Meier (1958). It can be used to measure the fraction of patients that survive for a certain amount of time after taking medication or receiving treatment, or the length of unemployment after people lose their jobs. I briefly introduce the Kaplan-Meier estimator below.

Assume that a total number of k events are observed for a sample of n individuals, where $k \leq n$. Assume these k events occur at distinct times, and $0 < t_1 < t_2 < \dots < t_k$. So, t_i is a time when at least one event happened. Let d_i be the number of events that happened at time t_i and n_i the number of subjects that have not yet had an event or have not been censored at time t_i , where $1 \leq i \leq k$. The Kaplan-Meier estimator for the survivor function is defined as

$$\hat{S}(t) = \prod_{i: t_i \leq t} \left(\frac{n_i - d_i}{n_i} \right).$$

For $t < t_1$, $\hat{S}(t)$ is defined as 1. For $t > t_k$, $\hat{S}(t) = 0$ if there are no censored times greater than t_k . $\hat{S}(t)$ will be undefined if the censored times are greater than t_k . The Kaplan-Meier estimator changes value only at the observed survival time.

The standard error of the Kaplan-Meier estimator is most commonly estimated by Greenwood's formula:

$$\text{var}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{i: t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}.$$

The variance formula helps to construct the limits of the $(1 - \alpha)100\%$ confidence intervals at the specific time t of $\hat{S}(t)$, i.e.,

$$\hat{S}(t) \pm z_{\alpha/2} \cdot \sqrt{\text{var}[\hat{S}(t)]}.$$

This confidence interval is time dependent and does not apply to the entire sample. The limitation of the Kaplan-Meier estimator is that we cannot use it to estimate survival after controlling for covariates.

The life-table estimator uses an algorithm that is similar to the Kaplan-Meier estimator. It also counts the actual number of events from the longitudinal data. It is also known as the actuarial-table estimator. It could handle a large sample of survival data in which survival times are grouped into intervals.

Assume the survival times are grouped into $k + 1$ intervals $(0, t_1)$, (t_1, t_2) , $\dots$, (t_{k-1}, t_k) , (t_k, ∞) . Let d_i be the number of events, c_i be the number of censored cases, and n_i be the number of subjects at risk at the beginning of the time interval (t_i, t_{i+1}) , where $0 \leq i \leq k$, $t_0 = 0$, and $t_{k+1} = \infty$. For any $j = 1, \dots, k$, the life-table estimator for the survivor function is defined as

$$\hat{S}(t_j) = \prod_{i=0}^j \left(\frac{n_i - \frac{c_i}{2} - d_i}{n_i - \frac{c_i}{2}} \right).$$

The estimation of the variance of life-table estimator is similar to the application of Greenwood's formula for Kaplan-Meier estimator,

$$\text{var}[\widehat{S}(t_j)] = [\widehat{S}(t_j)]^2 \sum_{i=0}^j \frac{d_i}{(n_i - \frac{c_i}{2} - d_i)(n_i - \frac{c_i}{2})}.$$

82.1.1.2 Parametric Hazard Models

The nonparametric methods do not allow researchers to control for explanatory variables, but parametric hazard models do. Assume a vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ for the explanatory variables of an individual i . Both continuous variables and indicator variables for qualitative classifications can be explanatory variables in parametric hazard models. The vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ represents the coefficients to be estimated. We assume that a linear predictor $\boldsymbol{\beta}^T \mathbf{x}_i = \sum_{j=1}^p \beta_j x_{ij}$ affects the survival rate through some function $\psi(\boldsymbol{\beta}^T \mathbf{x})$. The most common parametric models are the accelerated failure time (AFT) models.

Let S_0 and λ_0 be the baseline survival function and the baseline hazard function. AFT models assume $S_i(t) = S_0(\psi t)$ and $\lambda_i(t) = \psi \lambda_0(\psi t)$ for all t . The logarithm of survival time is

$$\log T_i = \sum_{j=1}^p \beta_j x_{ij} + \sigma \epsilon_i = \boldsymbol{\beta}^T \mathbf{x}_i + \sigma \epsilon_i,$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ represents the vector of unknown regression parameters, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ is the vector for a total of p explanatory variables of an individual i , and σ is an unknown scale parameter. Different AFT models are named after the distribution of *survival time* T rather than the distribution of ϵ . These models are estimated using the maximum likelihood method regardless of the distribution of T .

I first briefly describe the exponential model, one of the commonly used AFT models. If T is assumed to follow the exponential distribution, its pdf function is

$$f(t) = \lambda \exp(-\lambda t).$$

The survival function is

$$S(t) = \exp(-\lambda t),$$

where $\lambda = \exp(-\boldsymbol{\beta}^T \mathbf{x})$, and the hazard function is assumed to be independent of survival time t , i.e.,

$$\lambda(t) = \lambda > 0.$$

An extension of the exponential model is the Weibull model. If T is assumed to follow a Weibull distribution, its pdf function is

$$f(t) = \lambda \gamma t^{\gamma-1} \exp[-\lambda t^\gamma].$$

The survival function is

$$S(t) = \exp[-\lambda t^\gamma],$$

where $\gamma = \frac{1}{\sigma} > 0$, and $\lambda = \exp(-\beta^T \mathbf{x}/\sigma) > 0$. When the Weibull shape parameter σ equals one, the Weibull model is equivalent to the exponential model.

The log-normal model is when T is assumed to follow a log-normal distribution (i.e., $\log T$ follows a normal distribution). The corresponding pdf function is

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left\{-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^2\right\},$$

The corresponding survivor function is

$$S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right),$$

where $\beta^T \mathbf{x} > 0$, $\sigma > 0$, and Φ is the cumulative distribution function for the normal distribution.

The generalized gamma model is when T is assumed to follow a three-parameter generalized gamma distribution. Its pdf is

$$f(t) = \frac{\gamma \tau^\theta}{\Gamma(\theta/\gamma)} t^{\theta-1} \exp[-(\tau t)^\gamma], \theta > 0, \tau > 0, \gamma > 0$$

where Γ is a complete gamma function. Both the Weibull model and log-normal model are nested within the generalized gamma model. For example, the generalized gamma distribution is also the Weibull distribution (when $\theta = \gamma$ and $\lambda = \tau^\gamma$).

When T follows a log-logistic distribution (i.e., $\log T$ follows a logistic distribution), we have the log-logistic model. The pdf function is

$$f(t) = \frac{\lambda \gamma t^{\gamma-1}}{(1 + \lambda t^\gamma)^2},$$

and the survival function is

$$S(t) = \frac{1}{1 + \lambda t^\gamma},$$

where $\gamma = 1/\sigma$ and $\lambda = \exp(-\beta^T \mathbf{x}/\sigma)$. The log-logistic model has no nesting relationship with any of the previously introduced models.

82.1.1.3 Semiparametric Hazard Models

Unlike the parametric methods, semiparametric models do not assume the distribution of the survival time T . The Cox proportional hazard model is a widely used semiparametric model. The model assumes $S_i(t) = [S_0(t)]^\psi$ and the hazard function is $\lambda_i(t) = \psi\lambda_0(t) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_i)$, where S_0 and λ_0 denote the baseline survival function and baseline hazard function. $\boldsymbol{\beta}$ is the vector of unknown regression parameters, $\mathbf{x}$ is the vector of explanatory variables, and λ_0 is an unspecified baseline hazard function. Because $\boldsymbol{\beta}^T \mathbf{x}_i$ is still used for estimating the regression parameters, the Cox proportional hazard model is semiparametric method.

The hazard ratio is a relationship between the instantaneous hazards of two groups. For example, the hazard ratio for individuals i and j is

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{\lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_j)} = \exp[\boldsymbol{\beta}^T (\mathbf{x}_i - \mathbf{x}_j)].$$

In the Cox proportional hazard model, the hazard ratio was often used to measure the effect of an intervention on the occurrence of events (i.e., the treatment vs. the placebo). This ratio does not require the knowledge of the baseline hazard function λ_0 due to cancelation – a major advantage of the Cox proportional hazard model.

Another advantage of the Cox proportional hazard model is the ability of the model to incorporate time-dependent covariates. The hazard ratio does not reflect a time unit of the study. Because the hazard rate is an instantaneous risk that an event happens at time t , the hazard ratio can be simply interpreted as the relationship between the instantaneous hazards in two groups over the whole study time period.

82.1.1.4 Logistic Regression

Although the survival time is often assumed to be continuous, it can be discrete. If the survival time can only be observed in certain intervals (such as in the number of years), the data are commonly referred to as discrete-time data. Let the probability of an individual i experiencing an event at time t be P_{it} ($P_{i0} = 1$).

The odds that an event occurs at time t for an individual i if it has not happened is

$$O_{it} = \frac{P_{it}}{1 - P_{it}}.$$

The proportional odds model assumes $O_i(t) = \psi O_0(t)$ for all t , where O_0 denotes the baseline odds function.

By incorporating the explanatory variables, the model then becomes

$$O_i(t) = O_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_i).$$

O_0 is an unspecified baseline odds function, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ is the vector for a total of p explanatory variables of an individual i , and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ is the vector of unknown regression parameters. It follows that $\boldsymbol{\beta}^T \mathbf{x}_i = \sum_{j=1}^p \beta_j x_{ij}$.

The dependent variable is the logarithm of the odds of the outcome, so the coefficients measure the change in the odds ratios. The proportional odds model can be estimated by a standard maximum likelihood method through the traditional logistical regression model. The outcome variable follows a binomial distribution. The advantage of logistic regression is that it is a widely used parametric method and easy to understand.

For any two individuals, i and j , the ratio of their odds can drop out of the baseline hazard function because they are cancelled out.

$$\frac{O_i(t)}{O_j(t)} = \frac{O_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{O_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_j)} = \exp[\boldsymbol{\beta}^T (\mathbf{x}_i - \mathbf{x}_j)]$$

Similar to the Cox proportional hazard model, the ratio of the odds does not depend on time t . Sometimes, the proportional odds model is also called the Cox proportional hazard model for discrete-time data.

The Cox model is estimated using the partial likelihood function. Assume that there are n independent individuals, and for each individual i , the data consist of three parts: the observed time Y_i , the event indicator δ_i , and explanatory variables vector $\mathbf{x}_i$. The partial likelihood function can be written as

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^n \left[\frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{\sum_{t_j \geq t_i} \exp(\boldsymbol{\beta}^T \mathbf{x}_j)} \right]^{\delta_i},$$

where $\sum_{t_j \geq t_i} \exp(\boldsymbol{\beta}^T \mathbf{x}_j)$ includes all individuals who are still at risk at the survival time t_i . Accordingly, the log-partial likelihood is

$$pl(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i [\boldsymbol{\beta}^T \mathbf{x}_i - \log \sum_{t_j \geq t_i} \exp(\boldsymbol{\beta}^T \mathbf{x}_j)]$$

The score function and the Fisher information matrix function of the log-partial likelihood are

$$\begin{aligned} U(\boldsymbol{\beta}) &= \frac{\partial pl(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left[\delta_i \mathbf{x}_i - \frac{\sum_{t_j \geq t_i} \mathbf{x}_j \exp(\boldsymbol{\beta}^T \mathbf{x}_j)}{\sum_{t_j \geq t_i} \exp(\boldsymbol{\beta}^T \mathbf{x}_j)} \right], \text{ and } \\ I(\boldsymbol{\beta}) &= -\frac{\partial^2 pl(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \\ &= \sum_{i=1}^n \left[\frac{\sum_{t_j \geq t_i} \mathbf{x}_j \mathbf{x}_j^T \exp(\boldsymbol{\beta}^T \mathbf{x}_j)}{\sum_{t_j \geq t_i} \exp(\boldsymbol{\beta}^T \mathbf{x}_j)} - \frac{\sum_{t_j \geq t_i} \mathbf{x}_j \exp(\boldsymbol{\beta}^T \mathbf{x}_j) \sum_{t_j \geq t_i} \mathbf{x}_j^T \exp(\boldsymbol{\beta}^T \mathbf{x}_j)}{\left(\sum_{t_j \geq t_i} \exp(\boldsymbol{\beta}^T \mathbf{x}_j) \right)^2} \right]. \end{aligned}$$

Table 1 Comparing different models

	Parametric hazard model	Cox proportional hazard model	Logistic regression model
Model type	Parametric	Semiparametric	Parametric
Estimation method	Maximum likelihood method	Partial likelihood method	Maximum likelihood method
Dependent variable	The logarithm of the duration	The logarithm of the hazard	The logarithm of the odds of the outcome
Distribution of the dependent variable	It depends on the assumption of the survival time T	Not specified	The outcome follows the binomial distribution
Interpretation of the coefficient	Change in the duration	Change in the hazard ratios	Change in the odds ratios
Advantage	Provides efficient estimates if the shape of the distribution of T is known	No need for the baseline hazard or the distribution of T . Robust estimates	It can fit discrete-time data
Disadvantage	Need to specify the distribution of T correctly. Unable to incorporate time-dependent covariates	Usually appropriate for continuous-time data only	Not used for continuous-time data

Similar to parametric regression models, the Wald statistics, the score statistics, and the likelihood ratio statistics are common test statistics for semiparametric models.

If two or more events occur at the same time, the partial likelihood estimation might be invalid. Such events are quite common for data with annual or even monthly frequencies. The usual remedies to handle tied survival time are Breslow's approximation and Efron's approximation. These methods usually take some computational time because they are based on the exact method. An alternative solution is to consider the survival time as discrete and estimate it using logistic regression.

Another advantage of the Cox proportional model is that it can fit time-dependent covariates. Sometimes, all covariates may be time independent, which means they are covariates measured at the baseline ($t = 0$). It is also possible that covariates immediately prior to negative events are of more interest to researchers. Further, some covariates can vary by year, which means they can also be time dependent. The Cox proportional hazard model can be used to accommodate time-dependent covariates.

In Table 1, I present a table briefly summarizing and comparing these models. Parametric hazard models and logistic regression models are both parametric methods, whereas Cox proportional hazard model is a semiparametric method.

The parametric models are estimated using the maximum likelihood method, and Cox proportional hazard model is estimated using the partial likelihood method.

The dependent variables differ across these models. Regardless of the assumption of the distribution of the survival time T , the dependent variables are the logarithm of the duration until an event occurs. The dependent variable for the Cox proportional hazard model is the logarithm of the hazard. If the sign of a coefficient in a parametric regression model is positive, it would be negative in the Cox proportional hazards model. This is intuitive – if an explanatory variable increases the duration until an event occurs, it lowers the hazard or risk of the event. The dependent variable in the logistic regression model is the logarithm of the odds of the event. The expected sign of the coefficients in the logistic regression model is the same as those in Cox proportional hazards model.

The distribution of the dependent variable for parametric hazard models depends on the assumption of the survival time T . Such an assumption is not necessary for Cox proportional method. The distribution of the outcome variable is the binomial distribution in the logistic regression model. The coefficient estimates from parametric hazard models measure the change in the duration for changes in the explanatory variables; in Cox proportional hazard models, they measure the change in the hazard ratios; in logistic regression models, they measure the change in the odds ratios.

If the assumption the distribution of survival time T is known, parametric hazard models provide efficient estimates of the coefficients. However, it requires researchers to correctly specify the distribution of T , and it cannot be used to incorporate time-dependent covariates. In contrast, Cox proportional hazard model does not require researchers to specify either the baseline hazard or the distribution of T . In addition, it provides more robust estimates than parametric hazard models. This model can be used to estimate continuous-time data, and it can fit time-dependent covariates. The logistic regression model is suitable for fitting discrete-time data. It is easy to understand, and the coefficients are easy to interpret.

82.2 Literature Review

The literature has used various models to predict the likelihood of certain events, such as corporate failure. For example, Ohlson (1980) predicts the likelihood of corporate bankruptcy within 1 year using variables including total liabilities, working capital, net income (all deflated by total assets), firm size, and the ratio of funds from operations divided by total liabilities. He also predicts bankruptcy within 2 years if a company did not fail within 1 year or the most recent 2 years.

Also forecasting bankruptcy, Shumway (2001) develops a formal hazard model that can be viewed as a discrete accelerate failure time model or a logit model using all firm-year observations. Using data over 31 years, he finds that some accounting ratios, such as working capital, retained earnings, and sales (all deflated by total assets), are poor predictors of bankruptcy. He also introduces several market-driven

variables that are good predictors in forecasting corporate failures, including market size, past stock returns, and standard deviation of stock returns.

Recent literature has begun using the Cox proportional hazard model in forecasting various events, including personal and corporate bankruptcy, personnel turnover, hedge fund disappearance, acquisitions, IPOs, mergers and acquisitions, loan default, loan prepayment, and regulatory action against failing financial institutions. Agarwal, Chomsisengphet, Liu, and Mielnicki (2005) use the Cox proportional model to assess the impact of state bankruptcy exemption levels on small business owners' decision to file for bankruptcy. They find that higher levels of bankruptcy exemptions levels increase the likelihood of bankruptcy. Bhattacharjee, Higson, Holly, and Kattuman (2009) also use the model to evaluate the impact of macroeconomic instability on two forms of firm that exits in the UK, i.e., bankruptcies or acquisitions. They find an opposing effect of macroeconomic instability – it increases the hazard of bankruptcy but decreases the hazard of acquisition. Using survey data of Russian industrial firms, Iwasaki (2014) applies the Cox proportional hazards model to examine the survival of Russian industrial firms and finds the following important factors, i.e., corporate governance, human resource, and the influence of the board on management.

Brown, Goetzmann, and Park (2002) use the Cox proportional hazard model to examine factors contributing to the failure of hedge funds. They find that the success of hedge funds depends on both absolute and relative performance, excess volatility, and on fund age. Brown, Cookson, and Heimer (2019) use a setup like the Cox proportional hazard model and find that better financial market development on Native American reservations leads to a higher likelihood of entering credit markets.

Jenter and Kanaan (2015) use the Cox proportional hazard model to predict the forced turnover of CEOs and find that the effect of poor industry performance is stronger than the effect of market performance. Khanna, Kim, and Lu (2015) use the Cox method to examine whether CEO connectedness affects how soon fraud is detected. They find that it takes longer to uncover fraud when a CEO has better connections with top executives and directors through past employment, education, or social organization. Maturana and Nickerson (2020) examine the impact of a teacher filing for personal bankruptcy on the likelihood of keeping her job as a public school teacher in the state of Texas in the following year. Leverty and Grace (2018) find causal evidence that regulators delay interventions against failing financial institutions before elections. Fjesme, Galpin, and Moore (2021) use the model to investigate the likelihood of rejected applicants of the London Stock Exchange between 1891 and 1911 and find them more likely to file for liquidation than successful applicants.

Gerakos, Lang, and Maffett (2013) have also used the accelerated failure time model to predict the time until the suspension of trading on a stock exchange together with a negative cumulative return over the 60 days prior to delisting, assuming that the failure rates follow log-normal distribution. They find that firms listed on London's Alternative Investment Market (a privately regulated exchanges) have lower postlisting returns and failure rates than newly listed firms listed on traditionally regulated exchanges. These firms are comparable to those listed on US

Pink Sheets. Doidge et al. (2009) use the Cox proportional hazard model to predict the decision of a foreign firm to cross-list in the USA. Their findings suggest that controlling shareholders are less likely to cross-list when private benefits of controlling shareholders are high. Johnson (2005) also use the Cox model to predict the closure of investment accounts at mutual funds. They find observable shareholder characteristics that can be used to predict whether an account will be short term or long term on the day an account is opened. Becker, Ivkovic, and Weisbenner (2011) use the Cox proportional hazard model to predict when a stock is going to be sold after purchase, conditional on no prior sale. They find that a local senior investors is less likely to sell a share of stock than a nonlocal nonsenior investor.

Hazard models are often used in predicting the default of debt. For example, Banasik, Crook, and Thomas (1999) use it to estimate the length of time until default happens and thus examine when borrowers default rather than if they will default. This question is important because the lender could set the interest rate to compensate for the default risk. They use personal loan data from a UK financial institution. The data on some loans are censored due to the lack of default or prepayment as of the end of the sample period. They use a competing risk approach to model two likely events in the lifetime of a loan, i.e., prepayment of the loan or default. In other words, the survivor function models the distribution of both the length of time until the loan defaults and the time until the loan is repaid early.

In addition to predicting the default likelihood of individual borrowers, survival models are also used to study corporate bonds. For example, Moeller and Molina (2003) use the Cox proportional model to examine how the default probability of high-yield corporate bonds (i.e., junk bonds) changes over time. They use corporate bonds that appeared in the “Recent Corporate Bond Offerings” published by Moody’s with a rating of Ba1 or lower for approximately 6 years in the 1980s and find a constantly increasing risk of default over time, especially after the debt has been outstanding for 4 years.

Several researchers use different methods on the same data for comparison. Stepanova and Thomas (2002) use different extensions of the Cox proportional hazards model to analyze personal loan data in a major UK financial institution. Rather than using continuous characteristics such as age, they assign the values into 15–20 equal bands, thus developing a coarse classifying approach. They show that survival analysis models are competitive with (and sometimes superior to) the standard logistic model. Further, Bellotti and Crook (2009) show that the Cox proportional hazards model is competitive compared to logistic regression for predicting default on credit card accounts. They incorporate time-varying macro-economic variables such as bank interest rates, unemployment, and housing prices and show that these variables increase the model’s predictive ability. Lee and Kim (2014) find that the mixed Cox proportional hazards model predicts hedge fund failure more accurately than other models, such as fixed covariate proportional models and discrete logit models. Agarwal, Amromin, Ben-David, and Dinc (2018) use both parametric models (Exponential hazard, Weibull hazard, and linear probability model) and semiparametric models (Cox proportional hazard model) to predict the start of foreclosure in 2009–2010. They find that foreclosure is delayed

on delinquent mortgages in the districts of committee members by 0.5 months. The estimated costs to lenders is substantial. It is greater than the campaign contributions made by the political action committees of the largest mortgage servicing banks to the committee members.

82.3 Empirical Application

To illustrate the application of these methods, I obtain a sample of firms with class action lawsuits and examine the influence of employee whistleblowing on the probability of case dismissal. I identify 236 class action shareholder lawsuits filed in Federal courts from 2004–2014 from the Securities Class Action Services (SCAS) database using the Institutional Shareholder Services. Out of these lawsuits, 121 cases are dismissed during the sample. The rest are active lawsuits as of the end of the sample period. Firms with employee whistleblowing discrimination cases filed are based on data from the Occupational Safety and Health Administration (OSHA). Firm characteristics, such as the level and volatility of stock returns, total assets, and sales growth, come from COMPUSTAT and CRSP. Table 2 tabulates the details of the variable definition.

I start with parametric models. First, I use the logistic model to examine whether lawsuits against firms with employee whistleblowing are more or less likely to be dismissed, i.e.,

$$Prob(Dismissed_{t+k} = 1) = \beta_1 + \beta_2 * Whistleblowing_t + \beta_3 * Assets_t + \beta_4 * Sales_growth_t + \beta_5 * Return_t + \beta_6 * Volatility_t + \varepsilon \quad (1)$$

The SAS program is the following:

```
* 1. Parametric Estimation: Logistic;
```

Table 2 Variable definition

Variables	Definition
<i>Duration</i>	The number of years between the class action lawsuits and the date of dismissal for cases that are dismissed, or the end of the sample period, i.e., October 31, 2015
<i>Dismissed</i>	An indicator variable equal to 1 if a class action lawsuit is dismissed
<i>Whistleblowing</i>	An indicator variable equal to 1 if the firm has experienced a whistleblowing allegation and 0 otherwise
<i>Assets</i>	The natural logarithm of total assets
<i>Sales_growth</i>	The growth rate of annual sales in the past 3 years
<i>Return</i>	The firm's market-adjusted 12-month stock return prior to the lawsuit announcement
<i>Volatility</i>	The natural logarithm of the standard deviation of firm-specific returns in down weeks/the standard deviation of firm-specific returns in up weeks

Table 3 Estimating model (1) using logistic regression

Panel A Coefficient estimates				
Parameter	Estimate	Standard error	Wald Chi-Square	Pr > ChiSq
<i>Whistleblowing</i>	−1.6162	0.7168	5.0838	0.0242
<i>Assets</i>	0.2102	0.0860	5.9792	0.0145
<i>Sales_growth</i>	0.9166	0.5538	2.7392	0.0979
<i>Return</i>	−0.1977	0.2744	0.5193	0.4711
<i>Volatility</i>	−0.1527	0.2213	0.4761	0.4902
Panel B Odds ratio estimates				
Odds Ratio Estimates				
Effect	Point estimate	95% Wald confidence limits		
<i>Whistleblowing</i>	0.199	0.049 0.810		
<i>Assets</i>	1.234	1.043 1.460		
<i>Sales_growth</i>	2.501	0.845 7.405		
<i>Return</i>	0.821	0.479 1.405		
<i>Volatility</i>	0.858	0.556 1.325		

```

proc logistic data=firms descending;
model dismissed = Whistleblowing Assets Sales_growth Return
Volatility;
quit;

```

I summarize the results of estimating model (1) using the logistic regression in Table 3. Panel A tabulates the coefficient estimates, standard error, and p-values of the Wald Chi-square statistic. The variables that are significantly different from zero include *Whistleblowing*, *Assets*, and *Sales_growth*. For example, lawsuits associated with employee whistleblowing are less likely to be dismissed (coef. = -1.6162 ; $p\text{-value} = 0.0242$), while cases against firms with more assets and faster sales growth are more likely to be dismissed (coef. = 0.2102 , $p\text{-value} = 0.0145$; coef. = 0.9166 , $p\text{-value} = 0.0979$, respectively).

Panel B tabulates the odds ratio of the coefficients to evaluate the economic significance of these variables. The odds ratios are the exponentiated coefficients. For example, Panel A shows that the coefficient estimate of *Whistleblowing* is -1.6162 , its exponential is $\exp(-1.6162) = 0.199$ after rounding. Compared with lawsuits that are not associated with employee whistleblowing, the odds of a lawsuit being dismissed versus not dismissed are reduced by a factor of 0.199. This effect appears to be economically significant.

We next use parametric models with different assumptions of the distribution of T to estimate the duration (i.e., survival time) until an event (i.e., a case being dismissed) occurs. In this case, the longer the survival time, the less likely a lawsuit is dismissed. Depending on the assumption of the distribution of the survival time (i.e., exponential, Weibull, log normal, or log logistic), we choose one of the following SAS commands.

```

* 2. Parametric Estimation: Exponential;
proc lifereg data=firms;
  model duration*dismissed(0)=
Whistleblowing Assets Sales_growth Return Volatility /
D=exponential;
run;

* 3. Parametric Estimation: Weibull;
proc lifereg data=firms;
  model duration*dismissed(0)=
Whistleblowing Assets Sales_growth Return Volatility/ D=weibull;
run;

* 4. Parametric Estimation: Log normal;
proc lifereg data=firms;
  model duration*dismissed(0)=
Whistleblowing Assets Sales_growth Return Volatility/ D=lnormal;
run;

* 5. Parametric Estimation: Loglogistic;
proc lifereg data=firms;

```

Table 4 Coefficient estimates from parametric duration analysis

Panel A Exponential model		
Variables	Coef. estimate	P-value
<i>Whistleblowing</i>	1.6532	0.0056
<i>Assets</i>	−0.1329	0.0128
<i>Sales_growth</i>	−0.4875	0.1588
<i>Return</i>	0.0478	0.7881
<i>Volatility</i>	0.1037	0.4712
Panel B Weibull model		
Parameter	Coef. estimate	Estimate
<i>Whistleblowing</i>	2.2743	0.0175
<i>Assets</i>	−0.2039	0.0210
<i>Sales_growth</i>	−0.7375	0.1831
<i>Return</i>	0.1199	0.6827
<i>Volatility</i>	0.1369	0.5582
Panel C Log-normal model		
Parameter	Coef. estimate	Estimate
<i>Whistleblowing</i>	1.7135	0.0108
<i>Assets</i>	−0.1751	0.0270
<i>Sales_growth</i>	−0.6116	0.2115
<i>Return</i>	0.0605	0.8156
<i>Volatility</i>	0.0956	0.6415
Panel D Log-logistic model		
Parameter	Coef. estimate	Estimate
<i>Whistleblowing</i>	2.1642	0.0069
<i>Assets</i>	−0.2042	0.0149
<i>Sales_growth</i>	−0.7292	0.1566
<i>Return</i>	0.1093	0.7000
<i>Volatility</i>	0.1220	0.5749

```
model duration*dismissed(0)=  
Whistleblowing Assets Sales_growth Return Volatility/ D=Llogistic;  
run;
```

After estimating the four parametric duration models, we report the coefficient estimates for each of the four models in Table 4. Despite different assumptions of the distribution, the coefficient estimates have the same signs, suggesting the same inferences across these models. For example, the coefficient on *Whistleblowing* is positive and significantly different from zero in all models, mostly at the 1% level. This indicates that lawsuits against firms with employee whistleblowing have a longer expected duration and are less likely to be dismissed. Note that even though the coefficients have the opposite signs from the logistic model, the inference remains the same.

We need to transform the parameter estimates to obtain the economic significance. Take the coefficient on *Whistleblowing* from Panel A of Table 4 for an example – $\exp(1.6532) = 5.223$. Therefore, cases against firms with employee

Table 5 Coefficient estimates from semiparametric survival analysis

Parameter	Parameter estimate	Standard error	Chi-square	Pr > ChiSq	Hazard ratio
<i>Whistleblowing</i>	-1.2002	0.5959	4.0575	0.0440	0.301
<i>Assets</i>	0.1336	0.0565	5.5917	0.0180	1.143
<i>Sales_growth</i>	0.5805	0.3459	2.8147	0.0934	1.787
<i>Return</i>	-0.1500	0.1898	0.6244	0.4294	0.861
<i>Volatility</i>	-0.1071	0.1490	0.5168	0.4722	0.898

whistleblowing have an expected duration of 5.223 times that for cases without employee whistleblowing. The effect of whistleblowing on the likelihood of class action lawsuits appears economically significant, as the empirical findings based on logistic regression. Similarly, the inferences on *Assets* are also consistent for both logistic regression and parametric duration analysis. However, none of the coefficients on *Sales_growth* is significantly different from zero, which differs from the variable being marginally significant using logistic regression.

Next, we use a semiparametric method, i.e., the Cox proportional hazards model, to estimate the hazard function. The SAS program below illustrates how we implement the estimation, dealing with ties using Efron's approximation. We tabulate the coefficient estimates in Table 5. The statistical significance of the variables is similar to that using logistic regression.

However, the signs of the coefficients are exactly the opposite of those from the parametric duration analysis. Still, the inferences are the same – the Cox proportional hazard model estimates the hazard ratio, which indicates the occurrence of dismissal. In contrast, the parametric duration models estimate the survival time until dismissal. The economic significance of the coefficients can also be obtained by taking the exponential of the coefficient. For example, $\exp(-1.2002) = 0.301$. This means that the hazard of a lawsuit being dismissed decreases by 69.9% if the lawsuit is associated with employee whistleblowing.

```
* 6. Semiparametric Estimation: Cox Model;
proc phreg data=firms;
  model duration*Dismissed(0)=
  Whistleblowing Assets Sales_growth Return Volatility / ties=efron;
run;
```

82.4 Conclusion

This chapter briefly discusses a few commonly used hazard models in accounting and finance research. Researchers in these fields are often interested in predicting the time until bankruptcy, the decision to cancel intimal public offering, being delisted from an exchange, default of corporate debt and personal loans, and a dramatic

change in stock price. Hazard models can also be used to predict time until events such as executive turnover.

I first introduce parametric models based on the different assumptions of the distribution of the survival time. Next, I describe a semiparametric regression model, i.e., the Cox proportional hazards model. I also introduce the use of logistic regression when dealing with discrete-time data. Using data on class action lawsuits and employee whistleblowing data, I illustrate how to apply these methods in SAS and compare the inferences.

References

- Agarwal, S., S. Chomsisengphet, C. Liu, and L. Mielnicki. 2005. Impact of state exemption laws on small business bankruptcy decision. *Southern Economic Journal* 71: 620–635.
- Agarwal, S., G. Amromin, I. Ben-David, and S. Dinc. 2018. The politics of foreclosures. *Journal of Finance* 73: 2677–2717.
- Audrino, F., A. Kostrov, and J. Ortega. 2019. Predicting U.S. Bank Failures with MIDAS Logit Models. *Journal of Financial and Quantitative Analysis* 54(6): 2575–2603.
- Banasik, J.N., J.N. Crook, and L.C. Thomas. 1999. Not if but when will borrowers default. *The Journal of the Operational Research Society* 50: 1185–1190.
- Becker, B., Z. Ivkovic, and S. Weisbenner. 2011. Local dividend clienteles. *Journal of Finance* 66: 655–683.
- Bellott, A., and J. Crook. 2009. Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society* 60(12): 1699–1707.
- Bhattacharjee, A., C. Higson, S. Holly, and P. Kattuman. 2009. Macroeconomic instability and business exit: Determinants of failures and acquisitions of U.K. firms. *Economica* 76: 108–131.
- Brown, S.J., W.N. Goetzmann, and J. Park. 2002. Careers and survival: Competition and risk in the hedge fund and CTA industry. *Journal of Finance* 56: 1869–1886.
- Brown, J.R., A. Cookson, and R.Z. Heimer. 2019. Growing up without finance. *Journal of Financial Economics* 134: 591–616.
- Chen, R., R. Guo, and M. Lin. 2010. Self-selectivity in firm's decision to withdraw IPO: Bayesian inference for hazard models of bankruptcy with feedback. *Journal of the American Statistical Association* 105: 1297–1309.
- Doidge, C., G.A. Karolyi, K.V. Lin, D.P. Miller, and R.M. Stultz. 2009. Private benefits of control, ownership, and the cross-listing decision. *Journal of Finance* 64: 425–466.
- Fjesme, S.L., N.E. Galpin, and L. Moore. 2021. Rejected stock exchange applicants. *Journal of Financial Economics* 139: 502–521.
- Gao, F., and X. He. 2020. Chapter 120: Survival analysis: Theory and applications in finance. In *Handbook of financial econometrics, mathematics, statistics, and machine learning*, 4149–4208. Singapore: World Scientific Publishing Company.
- Gerakos, J., M. Lang, and M. Maffett. 2013. Post-listing performance and private sector regulation: The experience of London's alternative investment market. *Journal of Accounting and Economics* 56: 189–215.
- Iwasaki, Ichiro. 2014. Global financial crisis, corporate governance, and firm survival: The Russian experience. *Journal of Comparative Economics* 42: 178–211.
- Jenter, D., and F. Kanaan. 2015. CEO turnover and relative performance evaluation. *Journal of Finance* 70: 2155–2184.
- Johnson, W. 2005. Preditable investment horizons and wealth transfers among mutual fund shareholders. *Journal of Finance* 59: 1979–2012.
- Kaplan, E.L., and P. Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53: 457–481.

- Khanna, V., E.H. Kim, and Y. Lu. 2015. CEO connectedness and corporate fraud. *Journal of Finance* 70: 1203–1252.
- Lee, H.S., and T.Y. Kim. 2014. Dynamic prediction of hedge fund survival in crisis-prone financial markets. *Journal of Banking & Finance* 39: 57–67.
- Leverty, J.T., and M.F. Grace. 2018. Do elections delay regulatory action? *Journal of Financial Economics* 130: 409–427.
- Maturana, G., and J. Nickerson. 2020. Real effects of workers' financial distress: Evidence from teacher spillovers. *Journal of Financial Economics* 136: 137–151.
- Mills, John. 1926. Table of the ratio: Area to bounding ordinate, for any portion of Normal curve. *Biometrika* 18: 395–400.
- Moeller, T., and C. Molina. 2003. Survival and default of original issue high-yield bonds. *Financial Management* 32(1): 83–107.
- Ohlson, J.S. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18: 109–131.
- Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business* 74: 1–16.
- Stepanova, M., and L. Thomas. 2002. Survival analysis methods for personal loan data. *Operations Research* 50: 277–289.

Cube Root Utility Theory

83

Jack Clark Francis

Abstract

The von Neumann-Morgenstern (vNM) utility function suggested here has loss aversion and consistently increasing utility of wealth between $-\infty$ and ∞ . It is also concave over favorable outcomes and convex over unfavorable outcomes. The function contains a consistent preference for positively skewed outcomes. The weighted probabilities in Kahneman and Tversky's (KT's) cumulative prospect theory do not align with the vNM requirement that the probabilities sum to one; nevertheless, important advantages are gained from the vNM assumption.

Keywords

Utility function · Kahneman-Tversky prospect theory · Cube root function · von Neumann-Morgenstern (vNM) utility theory · Radical function · Risk aversion · Loss aversion · Absolute risk aversion · Relative risk aversion

83.1 Introduction

A cube root is a number x that satisfies the relationship $y = \sqrt[3]{x} = x^{1/3}$. Figure 1 illustrates two cube root utility functions that are mirror images of each other. Equation (1) represents an S-shaped solid curve in Fig. 1 that slopes continually upward. Economic theory assumes people always prefer more money to less money; the solid curve in Fig. 1 reflects that assumption. The dotted line in Fig. 1 that decreases monotonically does not

Professor Francis is indebted to the Bert and Sandra Wasserman Endowment for support. Barry Ma, Norman Kleinberg, Arie Harel, Giora Harpaz and Jonathan Fung provided valuable assistance.

J. C. Francis (✉)

Economics and Finance, Baruch College, CUNY, New York, NY, USA

e-mail: jack@jcfrancis.com

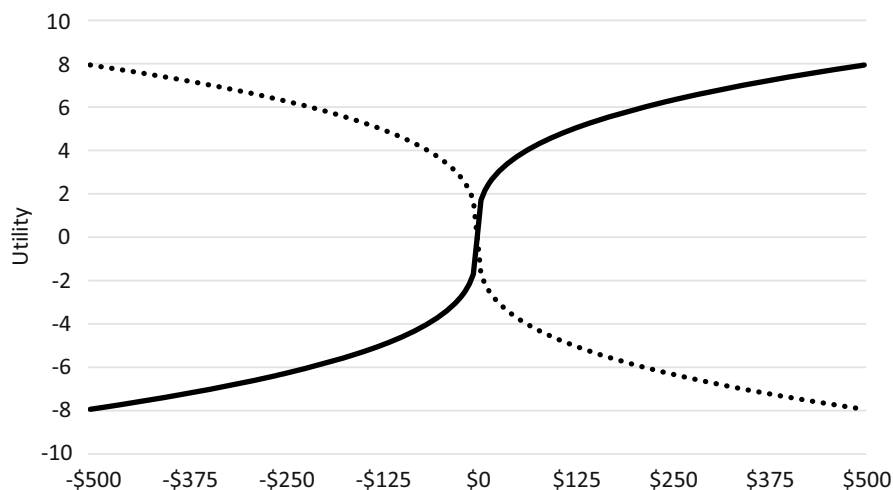


Fig. 1 Two symmetric cube root gain-loss functions

Note: The two cube root utility functions in Fig. 1 are mirror images of each other, but the dotted curve is implausible and is not be discussed further

represent plausible economics because it illustrates the preferences of someone who always prefers less money over more money. Therefore, this entry proposes and analyzes the solid cube root utility function in Fig. 1 that is defined in Eq. (1).

$$U(x) = \sqrt[3]{x} = x^{1/3} \quad dU/dx > 0 \quad (1)$$

Figure 2 is a modified version of the solid curve in Fig. 1. The solid curves in Figs. 1 and 2 both illustrate cube root utility functions that manifest the preferences suggested by Daniel Kahneman and Amos Tversky's (1979, 1992) prospect theory.¹

83.1.1 Initial Analysis

This chapter suggests a von Neumann-Morgenstern (vNM) cube root utility function, $U(x)$, which generates all the behaviors in Kahneman and Tversky's prospect theory (1979, 1992) except the weighted probabilities. This vNM theory is defined on dollar gains and losses measured from the origin and these gains and losses are represented by x in the radical Eq. (1).

Although the economics in Eq. (1) mimics the economics in the KT's prospect theory, the mathematics in the cube root utility function are more useful because

¹See Kahneman and Tversky (1979). Also see Tversky and Kahneman (1992). These articles and related articles are reproduced in a book of readings edited by Kahneman and Tversky (2000).

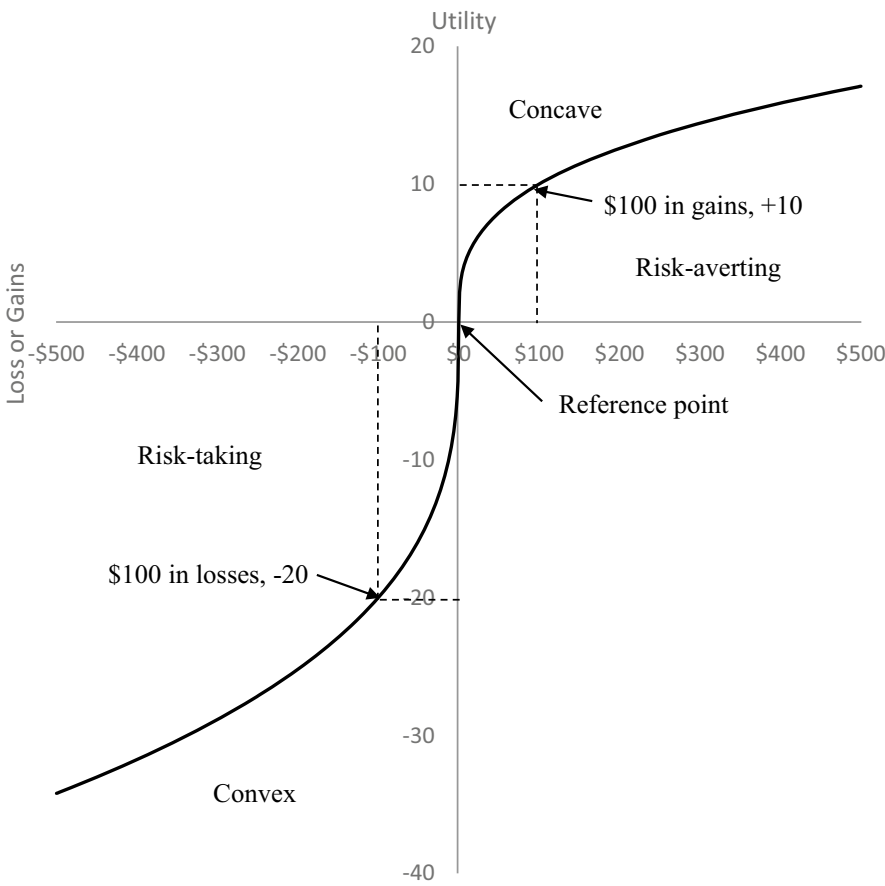


Fig. 2 A cube root utility function that aligns with Kahneman and Tversky
Notes: The cube root utility function is a reconfigured version of the solid S-shaped curve in Fig. 1. It is defined by the two following cube root equations: $y = 2\sqrt[3]{10} x^{1/3}, x < 0$, and $y = \sqrt[3]{10} x^{1/3}, x > 0$

vNM utility theory requires that the probabilities sum to one. As a result, some impressive behavioral research supporting KT’s cumulative prospect theory model with weighted probabilities is not included here.² What is more essential here is the mathematical expectation, standard deviation, variance-covariance matrix, and other important classical probability theories that can be used only if the probabilities sum to one, as assumed in vNM utility theory. Except for the fact that weighted probabilities are not used in the vNM model proposed here, all the rest of KT’s prospect

²Kahneman and Tversky (KT) introduced prospect theory with weighted probabilities in 1979 and 1992. This branch of behavioral economics is important and research in that direction should continue in the future. A paper by J.C. Francis that is similar to this chapter is published in the Review of Quantitative Finance and Accounting, 2021.

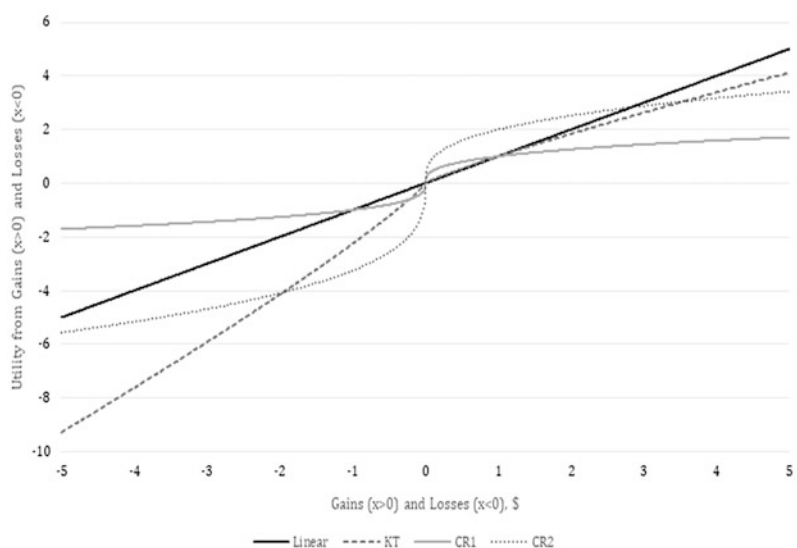


Fig. 3 Three S-Shaped Functions and One Linear Function, $U(x)$

Notes: The straight black line through the origin in Fig. 3 is a reference line representing a linear utility function with a slope of $dU(x)/dx = 1$. The Kahneman-Tversky value function is the dashed curve denoted KT in the legend. The solid gray curve that is symmetric about the origin represents Eq. (1); this is the simplest cube root utility function. The dotted curve represents the cube root utility function with each negative util multiplied by 3.25 and each positive util multiplied by 2; these two simple modifications show how easy it is to bring the cube root function closer to KT's dashed curve (see footnote 5)

theory emerges unscathed. Reformulating KT's prospect theory in a vNM format creates a new model that is more compatible with many existing economic, finance, decision theory, and psychological models. In addition, it will be shown below that cube root utility theory also provides important advantages over other popular vNM utility functions such as the log function and the square root function. Instead of competing with or detracting from KT's prospect theory, the vNM model introduced here suggests new and useful ways to employ prospect theory.

The linear utility function (black line), which has a slope of $dU/dx = 1$ in Fig. 3, is a reference line. Figure 3 compares Eq. (1) (the gray curve, denoted CR1 in the legend) with the Kahneman-Tversky value function (dashed curve, denoted KT in legend). KT have referred to graphs like these two curves as "S-shaped functions" (Kahneman and Tversky 1979, page 179).³ Generally over losses KT's S-shaped

³We follow KT's use of the phrase "S-shaped function" (KT, 1979, page 179). Their exponent of 0.88 is so close to 1.0 that KT's value curve is nearly linear. In fact, several researchers prefer to use piecewise linear approximations to our S-shaped function; for instance, see Nicholas Barberis and Wei Xiong, 2012. Unfortunately, piecewise linear functions are not able to span the space between positive infinity and negative infinity with a single equation. See Harel, Francis and Harpaz (2018) about functional forms.

function has a steeper slope than Eq. (1). This is because KT uses an exponent of +0.88 instead of an exponent of +1/3 for all x , and KT multiply the losses by a coefficient of 2.25 to compute the disutility from losses.⁴ The dotted curve (CR2 in legend) shows how simple it is to manipulate the cube root function to make it align more closely with KT's curve; this is accomplished by making the losses in Eq. (1) slope downward more steeply and by making the gains slope upward more steeply.⁵ Eq. (1) and KT's value function (that is, CR1 and KT in Fig. 3) are both S-shaped functions with a point of inflection at the origin. They can easily be modified to resemble each other more closely than illustrated in Fig. 3.

Equation (1) is represented by a single continuous line between positive infinity and negative infinity. It has both absolute risk aversion (ARA) and relative risk aversion (RRA), generates a consistent preference for positively skewed outcomes, and never conflicts with (Kahneman and Tversky 1979) or (Tversky and Kahneman 1992).⁶ The positively sloped cube root function in Eq. (1) is the simplest form of an infinite number of different S-shaped radical utility functions. Many, but not all, of these radical utility functions have desirable characteristics. In the conclusion, it will be shown below that the square root, logarithmic, exponential, and other frequently used utility functions do not possess as many desirable characteristics as the cube root utility function. Furthermore, each involves significant deficiencies.

Section 83.1.2 contains a review of the literature. Every research item discussed supports KT's prospect theory, the vNM cube root utility function, and the other vNM qualified radical utility functions because they are all economically similar. Section 83.2 dichotomizes various utility functions into two groups: functions that are qualified to analyze utility theory and functions that are unqualified. Section 83.3 explains why the cube root utility function is a useful utility function. Section 83.4 contains an initial analysis of the cube root utility function. Section 83.5 focuses on the gains in the concave segment (the positive line segment) of the cube root utility function. Section 83.6 analyzes the losses in the convex segment (the negative line segment) of the cube root utility function. Section 83.7 uses the Taylor Expansion to analyze the decision-maker's skewness

⁴KT's cumulative prospect theory paper (1992, Eq. (5)) analyzes the following S-shaped function:

$$U(x) = \begin{cases} x^{0.88}, & x \geq 0 \\ -2.25(-x)^{0.88}, & x < 0 \end{cases}$$

This is the equation underlying the dashed curve denoted KT in Fig. 3.

⁵The dotted curve in Fig. 3 represent: $U(x) = \begin{cases} 2x^{1/3}, & x \geq 0 \\ -3.25(-x)^{1/3}, & x < 0 \end{cases}$

⁶See Kahneman and Tversky (1979) and Tversky and Kahneman (1992).

preferences. Section 83.8 contains summary and concluding remarks. Appendix A generalizes the cube root results discussed in this chapter to other vNM radical utility functions.

83.1.2 Review of the Literature

Framing a Utility Function Harry Markowitz (1952) was the first to analyze the utility from gains and the disutility from losses relative to a *reference point* at the origin⁷, as KT have acknowledged (KT 1979, page 276). Markowitz proposed a utility function that contained both concave and convex segments. However, Markowitz left it for KT to link the concave region of the utility function with gains and the convex region with losses.

The Disposition Effect In 1985 Hersh Shefrin and Meir Statman analyzed the tendency of investors to prematurely sell stocks while they were still appreciating (Shefrin and Statman 1985). They also reported that investors tend to hold depreciating stocks for too long. These two unprofitable investor tendencies are referred to as the “disposition effect.” In 1998, Terrance Odean reported that brokerage accounts of 10,000 investors he analyzed provided further evidence supporting the disposition effect (Odean 1998b). Andrea Frazzini (2006) analyzed quarterly market data covering 29,000 mutual funds and found additional support for the disposition effect (Frazzini 2006). Other researchers have also published findings that support the disposition effect; some of these are discussed below.

Prospect theory provides theoretical explanations for the empirically observed aspects of the disposition effect. Shefrin and Statman, Odean, and Frazzini studied investors’ decisions to sell appreciating stocks before they reached their peak prices. Some of these appreciating stock prices will rise up into the concave section of KT’s utility function and this will cause the investor to become risk-averse. This increasing risk-aversion motivates investors to liquidate the appreciating assets. Second, Shefrin, Statman, Odean, and Frazzini studied investors’ reluctance to sell stocks when their prices were declining. Declining stock prices move down into the convex section of KT’s utility function where risk-taking becomes attractive. The resulting increased willingness to assume risk makes the investor increasingly likely to hold a depreciating stock in hopes that its falling price will rebound. Third, these scholars studied empirical evidence that investors fear losses more than they enjoy equal-sized gains. When the declining price of an asset falls through the inflection point at the origin, KT’s value function changes from concave to steeply convex. This increase in prospect theory’s loss aversion function explains why investors fear losses more than they enjoy gains of equal size.

⁷Markowitz (1952). KT acknowledge Markowitz’s contributions near their Fig. 3 in their 1979 paper. Kahneman and Tversky (1979).

Frazzini (2006) also discusses how security prices can underreact to news. He argues that if many investors display the disposition effect, stock prices may “underreact” to news, returns may be more predictable, and post-announcement prices may drift.

Mental Accounting In 1985 Richard Thaler noted that many people categorize life’s events in economic terms. Some of these people create mental accounts that are retained and used in decision-making, self-control, budgeting expenditures, planning, and saving for specific purposes (such as saving for college, planning a trip, or medical care). Furthermore, many investors categorize profitable investments as good outcomes and unprofitable investments as bad outcomes (Thaler 1985).⁸ The way a person chooses to assign resources to such categories can unintentionally create biases and departures from rational decision-making. Once adopted, these departures from rationality can become repetitive and might even be lifelong. Understanding the shortcomings of mental accounting can improve decision-making and open new research opportunities.

Grinblatt and Han (2005) use mental accounting and prospect theory to create a profitable momentum trading strategy (Grinblatt and Han 2005). (Frydman et al. 2018a, b) extended Thaler’s mental accounting to include what they call rolling mental accounting (Frydman et al. 2018a, b). Their laboratory experiments show that reinvestment decisions can cause reductions in the disposition effect that increase the profitability of trading.

The Mehra-Prescott Equity Risk Premium Puzzle In 1985 Rajnish Mehra and Edward Prescott reported a surprisingly large risk premium between the rates of return from common stock and Treasury bills during the century before 1985 (Mehra and Prescott 1985). They said it is difficult to explain this large empirical risk premium because the amount of risk aversion needed to justify it is theoretically implausible. Many empirical studies tried and failed to rationalize the oversized Mehra-Prescott equity risk premium. Then, in 1995 Shlomo Benartzi and Richard Thaler offered a two-part theoretical explanation for the over-sized Mehra-Prescott equity risk premium (Benartzi and Thaler 1995). First, Benartzi and Thaler suggested that investors are very loss averse. In other words, most investors are much more afraid of losses than they are pleased by an equal-size gain. These asymmetric risk preferences were previously explained by KT’s prospect theory (1979). Second, Benartzi and Thaler report that investors act as if they evaluate their portfolios annually, even if the portfolio’s planning horizon spans decades. That is, investors make their portfolio decisions as if they were operating with a one-year planning horizon. This myopic behavior is called “narrow framing.”

⁸ Also, see Barberis and Huang 2006.

Narrow Framing Barberis et al. (2006) provided further support for Benartzi and Thaler's 1995 finding that investors make decisions as if they were operating with a 1-year planning horizon (Barberis et al. 2006). The decision-making process called "narrow framing" gained popularity. Since 1995, much important research that is discussed below uses narrow framing to enrich the research. Recently, for example, Nicholas Barberis, Lawrence Jin, and Baolian Wang (October 2019) used narrow framing to investigate 22 prominent stock market anomalies.⁹

Realization utility As mentioned above, Richard Thaler's research into mental accounting revealed that many investors categorize profitable investments as good outcomes and unprofitable investments as bad outcomes. This focus on realized gains and realized losses motivated Nicholas Barberis and Wei Xiong to develop a utility realization theory based on an individual's investment transactions (Barberis and Xiong 2012). Realization utility is a narrowly framed analysis of decision-making that focuses on transactions involving a single stock, house, or mutual fund. The Barberis-Xiong realization utility theory model can rationalize the disposition effect and other behaviors that are part of prospect theory.¹⁰

Investigating narrow framing In 2001 Barberis and Huang created an asset pricing model based on mental accounting, loss aversion, the "house money effect," and several other realistic market assumptions (Barberis and Huang 2001a, b). The model simulates investor 1-year forecasts and decisions involving 500 stocks over 1–10,000 time periods (years). These experimental results are used to explore narrow framing and other implications of the model. Barberis and Huang's computer simulations support prospective theory and narrow framing. The next paragraph combines the Barberis and Huang conclusions with other important empirical evidence.

Empirical Investigation of Behavioral Finance Theory In 2016 Barberis, Mukherjee, and Wang assessed prospect theory, mental accounting, loss aversion, and narrow framing empirically using *historical* returns from individual stocks.¹¹ Their discussion of (their Fig. 1) explains how historical relative frequencies are transformed into weighted probabilities. Both the Barberis, Mukherjee, and Wang 2016 historical study and the 2001 Barberis and Huang forecasting model that was mentioned above develop results that are supportive of prospect theory. In their conclusions about these two studies, Barberis, Mukherjee, and Wang hypothesize

⁹Nicholas Barberis, Lawrence Jin, and Baolian Wang. "Prospect Theory and Stock Market Anomalies," NBER Working Paper 27155, May 2020, DOI 10.3386/w27155.

¹⁰The Barberis-Xiong realization utility theory differs from the vNM utility model analyzed here. Furthermore, Barberis and Xiong focus on utility functions that are linear or piecewise linear, which differs from this entry's curvilinear utility theory.

¹¹Barberis, Nicholas and Mukherjee, Abhiroop and Wang, Baolian, 2016. Prospect Theory and Stock Returns: An Empirical Test. February Issue. *Review of Financial Studies*, see Section "4 Conclusions."

that “sophisticated investors may exert effort to try to forecast a stock’s future return distribution; less sophisticated investors, on the other hand, may content themselves with thinking about a stock in terms of its past returns” (2016, Section 4 Conclusions).

Analyzing KT’s S-shaped value function Li and Yang built an equilibrium model to analyze prospect theory’s S-shaped value function (Li and Yang 2013). Under numerical parameters, which align with prospect theory, they show four findings:

- (i) The concavity/convexity of the value function can drive the disposition effect.
- (ii) The disposition effect can lead to momentum in stock returns that generates annual price momentum of roughly 1%.
- (iii) The disposition effect can explain why there is more trading in rising markets than in falling markets.
- (iv) The concavity/convexity of the value function alone, in the absence of loss aversion, raises equity premiums. These findings support other prospect theory studies by different researchers.

Empirical Price Illusion Birru and Wang provide empirical evidence that investors suffer from a nominal price illusion that causes them to overestimate the room for price growth in low-priced stocks relative to the room for price growth in high-priced stocks (Birru and Wang 2016). They report widespread evidence that investors systematically overestimate the skewness available from investments in low-priced stocks.

Empirical Study of Jackpot Payoffs Conrad, Kapadia, and Xing show that firms with a high potential for default also tend to have a relatively high probability of achieving large jackpot payoffs (Conrad et al. 2014a, b). This result is consistent with the well-known investor preference for skewed, lottery-like payoffs.

Empirical Analysis of Loss Realizations In 2016 Alex Imas demonstrated that following a realized loss, individuals tend to avoid risk. In contrast, he also showed that if the loss remains unrealized, most people tend to take on greater risk (Imas 2016).

Weighted Probabilities Weighted probabilities are an important part of prospect theory. Weighted probabilities should continue to be developed, but not with the vNM model suggested here. Consider a simple risky choice to clarify these issues. Many people would be willing to pay \$10 for an honest gamble with a one-in-ten chance to win \$100. According to vNM utility theory, this preference is irrational because the expected utility from winning, $E[U(0.1 \times \$100)]$, exactly equals the absolute value of the utility from the entry fee, $|U(-\$10)|$. It is simply not logical to pay \$10 to enter a gamble that does not increase the gambler’s expected utility. But most people overlook this rational vNM explanation and make their decision by overweighting the small probability of the large gain, as follows: $E[U(0.11 \times \$100)]$

$> |U(-\$10)|$. The 11% weighted probability in this decision is determined psychologically; it is not an objective probability. This decision violates the axioms of vNM utility theory and the foundations of classical probability theory. However, KT and other behavioral economists have conducted numerous experiments with weighted probabilities and have published impressive laboratory evidence that this is the way many people actually think.¹²

Separating Two Behavioral Economics Models KT's Eq. (6) in their cumulative prospect theory paper (1992), which is illustrated in KT's Figs. 1 and 2, create a transfer function between KT's weighted probabilities and the vNM objective probabilities. In their 2016 study, Barberis, Mukherjee, and Wang provide details about these computations in their Fig. 1 and its accompanying discussion.¹³ These complicated computations can be avoided by using the popular vNM model suggested here. Using vNM probabilities that sum to one is suggested for three reasons. First, using vNM probabilities permits using the mathematical expectation, standard deviation, variance-covariance matrix, and other powerful aspects of classical probability theory that cannot be used with KT's weighted probabilities that do not sum to one. Second, using the vNM model is more likely to generate cross-fertilizing research between economics, finance, decision theory, and psychology because these theories are mathematically compatible with the objective vNM probabilities. Third, Markowitz's Nobel Prize winning portfolio theory (Markowitz 1952), Sharpe's Nobel Prize winning Capital Asset Pricing Model (Sharpe 1964), Black-Scholes Nobel Prize winning option pricing theory (Black and Scholes 1973), and other important theories are based on objective vNM probabilities that are not compatible with KT's weighted probabilities. KT's weighted probabilities are incompatible with the Nobel Prize winning theories and that incompatibility hinders research efforts to compare and connect these important theories.

The preceding review of the literature provides strong support for KT's prospect theory and weighted probabilities.¹⁴ Research into behavioral economics with weighted probabilities is clearly important and should be continued. But we argue that the behavioral economics research with weighted probabilities should be continued separately from research involving the vNM model suggested here. Separating these two important behavioral economic models would not detract from either one, but it would make both more straightforward and easier to employ.

¹²See Daniel Kahneman and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*. 47(2): 263–91. Also see Amos Tversky and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*. 5(4): 297–323.

¹³Barberis, Nicholas and Mukherjee, Abhiroop and Wang, Baolian. 2016. "Prospect Theory and Stock Returns: An Empirical Test." February Issue. *Review of Financial Studies*.

¹⁴For additional review of prospect theory research findings see Nicholas Barberis 2013, "Thirty Years of Prospect Theory in Economics: A Review and Assessment," *Journal of Economic Perspectives*, Vol. 27, No. 1: 173–196.

The prospect theory research reviewed above motivated the development of the vNM cube root utility theory suggested here. The cube root utility function rationalizes the same economic issues that KT's cumulative prospect theory rationalizes, but the vNM model is based on the more widely used theories about utility and classical probabilities.

Before delving into the capabilities of cube root utility theory let us pause to investigate some complications that severely cripple several well-known utility functions.

83.2 Qualified and Unqualified Utility Functions

Log Utility Since logarithms are not defined for negative numbers, the log function cannot analyze losses. Thus, the positive utility curve $U(x) = \ln(1 + x)$, $x \geq 0$, is often accompanied by a negative *workaround function*, for instance, $-\ln(1 - x)$, $x < 0$, to create what appears to be, but is actually not, a single S-shaped function, which spans the horizontal axis between $-\infty$ to $+\infty$. Thus, the solid black curve in Fig. 4 is really composed of a convex function and a significantly different concave function that are linked at the origin to form a complicated S-shaped curve.

The Square Root Function Like the log function, the square root is severely handicapped because it is not defined over negative numbers. The negative workaround function: $U(x) = -(-x)^{1/2}$, $x < 0$, is sometimes added to the positive curve, $U(x) = x^{1/2}$, $x \geq 0$, to create what, falsely, appears to be one continuous gray S-shaped curve in Fig. 4. Actually, Fig. 4 illustrates two different concave functions that are linked at the origin to form a contrived S-shaped curve.

In contrast to the log and square root functions, the cube root function in Fig. 4 is represented by the one single, simple dashed curve, $U(x) = x^{1/3}$, which is defined in Eq. (1). Both the first and second derivatives of Eq. (1) are undefined at the origin, but we can still use a single equation between $-\infty$ to $+\infty$.

Summarizing, the solid black curve and the solid gray curve in Fig. 4 are assembled from a total of four different curvilinear functions. The log and square root workaround functions in Fig. 4 span the negative portion of the real line in an artificial and contrived manner. These workaround functions differ from the associated log and square root curves that underlie the positive half of the S-shaped utility curves.¹⁵ The negative and positive segments have different derivatives. These differences disqualify the log, square root, and similarly constructed utility functions from being able to analyze problems that involve any losses.

¹⁵The complete solid black log function: $U(x) = \begin{cases} \ln(1 + x), & x \geq 0 \\ -\ln(1 - x), & x < 0 \end{cases}$.

The complete solid gray square root function: $U(x) = \begin{cases} x^{1/2}, & x \geq 0 \\ -(-x)^{1/2}, & x < 0 \end{cases}$.

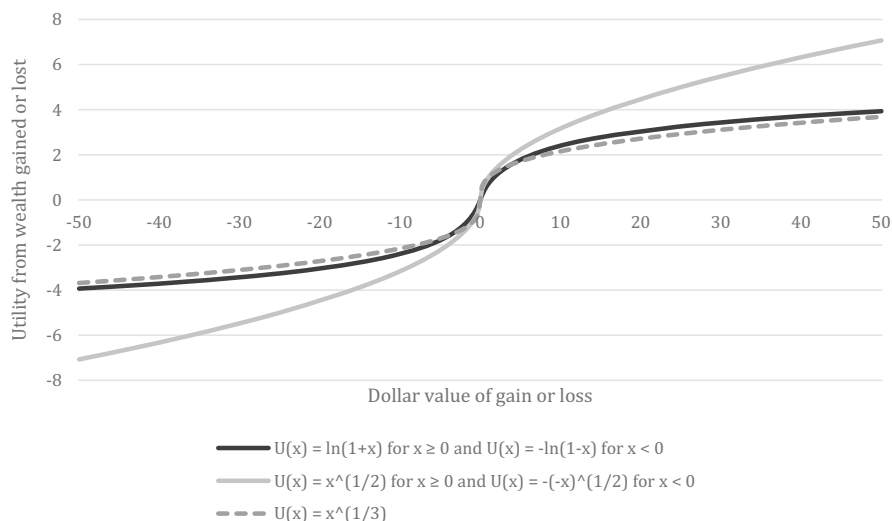


Fig. 4 Two unqualified and one qualified utility function

Unqualified Radical Functions The infinite number of functions of the form $y = x^{1/n}$ are called *radical equations*. The value of n is very important in radical functions. Numerous qualified radical equations with n an odd, positive integer, can span the real set with a single line. These radical functions are qualified to analyze losses, $x < 0$. For example, the cube root function and many similar radical utility functions with exponents like $\frac{1}{5}$, $\frac{1}{7}$, $\frac{1}{9}$, $\frac{1}{11}$, ad infinitum, are qualified to analyze losses.

The cube root function in Eq. (1) and Figs. 1, 2 and 3 is one of the qualified radical utility functions. In addition, other cube root utility functions can be constructed, for example, $U(x) = ax + bx^{1/3} + cx^{1/5}$. Various versions of the cube root utility functions can be formulated to be qualified utility functions.

In contrast, all radical utility functions with even integer values of n in $y = x^{1/n}$ are *unqualified* for further utility analysis because they are undefined for negative values of x . The square root function is a popular example of a radical utility function that is unqualified to analyze losses because $n = 2$ and 2 is not an odd integer.

83.3 Selecting the Most Desirable Radical Utility Function

Figure 5 illustrates three qualified radical utility functions that all generate similar economic behaviors.¹⁶ These three are a tiny fraction of the countably infinite number of similar qualified radical utility functions that could be illustrated in

¹⁶Appendix A analyzes these findings in more detail and suggests more general conclusions.

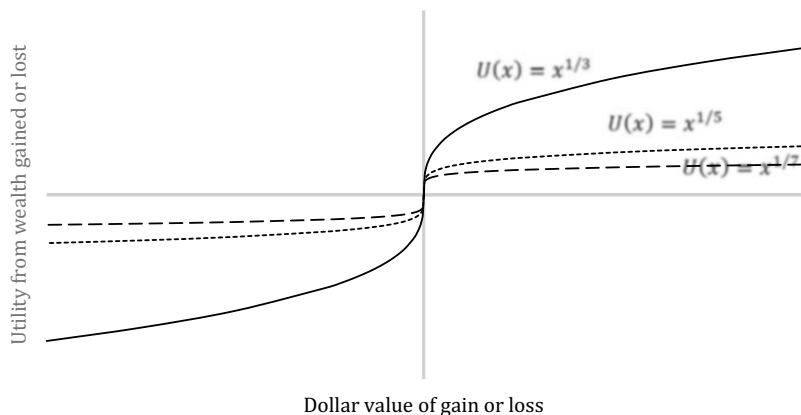


Fig. 5 Three qualified S-shaped utility functions

Fig. 5. Nevertheless, Fig. 5 suggests clearly that the cube root utility function generates a wider range of utility outcomes than other radical utility functions. A wider range of utility outcomes is more desirable than a smaller range of outcomes because it illustrates the results more clearly and dramatically. For this reason, the cube root utility function was selected as the most desirable radical utility function. As a result, the cube root utility function is the focus of this analysis. If circumstance arise in which a narrower range of responses is useful, a different radical utility function can be selected for that analysis.

Notes: If many qualified radical vNM utility functions were graphed together, they would asymptotically approach the horizontal axis as their exponents approached zero.¹⁷ The cube root radical utility function generates a wider range of utility outcomes than any other qualified radical utility function, which makes it more informative in many applications than the other qualified radical utility functions that are closer to the horizontal axis

83.4 Analysis of the Cube Root Utility Function

To reproduce Eq. (1) in different formats you can set $\alpha_1 = \alpha_2 = 1$ and $\gamma_1 = \gamma_2 = 1/3$ to derive Eqs.(3) and (4), for example.

$$U(x) = \sqrt[3]{x} = x^{\frac{1}{3}} \quad (2)$$

¹⁷To be more specific, the utility approaches +1 when x is positive, and -1 when x is negative.

$$= \begin{cases} \alpha_1 x^{\gamma_1}, x \geq 0, \alpha_1 = 1, \gamma_1 = 1/3 \\ \alpha_2 x^{\gamma_2}, x < 0, \alpha_2 = 1, \gamma_2 = 1/3 \end{cases} \quad (3)$$

$$= \begin{cases} x^{1/3} \\ -(-x)^{1/3} \end{cases} \quad (4)$$

The solid gray S-shaped curve denoted CR1 in Fig. 1 illustrates the cube root utility function that occurs when $\gamma_1 = \gamma_2 = 1/3$ and $\alpha_1 = \alpha_2 = 1$. The dotted S-shaped utility curve in Fig. 3 is an asymmetric cube root utility function with $\gamma_1 = \gamma_2 = 1/3$, $\alpha_1 = 2$ and $\alpha_2 = 3.25$. Following KT's (1992) numerical example, prospect theory is illustrated as the dashed curve in Fig. 3 with $\gamma_1 = \gamma_2 = 0.88$, $\alpha_1 = 1$ and $\alpha_2 = 2.25$.

Tversky and Kahneman (1981, 1992) and others have shown that people tend to avoid risk when a prospect is presented in a favorable environment, $x > 0$. That is, during a bull market most investors become risk-averse. However, many of these same investors might become risk-seekers when an equivalent outcome is introduced in a negative framework, $x < 0$. More specifically, during a bear market many people become risk-takers. This *cognitive bias* occurs as people react differently when asked to choose between positive and negative environments (two different frames) that have outcomes with the same absolute value. *Framing* helps reveal this cognitive bias.

Following Markowitz (1952), KT incorporated framing into utility theory by bifurcating the utility (value) function at the origin in their 1979 and 1992 papers. Figure 3 shows different qualified cube root utility functions that are bifurcated at the origin to form a concave (risk-averse, upper, happy frame) segment and a convex (risk-taking, lower, unhappy frame) segment. The concave and convex segments can be linked at the origin to form a single function. Eq. (3) can also represent asymmetric line segments when $\alpha_1 \neq \alpha_2$.

Barberis and Huang show that it is the kink at the origin in a utility function that creates *loss aversion* (Barberis and Huang 2001a,b). Qualified radical vNM utility functions like the cube root utility function in Eq. (1), was selected to be the focus of this chapter because it provides a single continuous line between positive infinity and negative infinity that can be kinked at the origin. The asymmetric dotted cube root utility function in Fig. 3 illustrates the case in which $\alpha_1 = 2$ and $\alpha_2 = 3.25$. In this case, the dotted utility function in the negative segment of Fig. 1 represents stronger loss aversion than the solid gray curve that has $\alpha_1 = \alpha_2 = 1$.

83.5 Analysis of the Utility Function with Favorable Outcomes, $x > 0$

The positive half of the solid gray S-shaped curve in Fig. 3 represents Eq. (3). This positive line segment is analyzed in this section, and the negative line segment is analyzed in Sect. 83.6. The γ_1 exponent in Eq. (3) is set equal to 1/3 to create Eq. (5).¹⁸

¹⁸Results like those reported in Sects. 83.5 and 83.5 can also be obtained with exponents of $\frac{1}{5}, \frac{1}{7}, \frac{1}{9}, \frac{1}{11}$, ad infinitum, as illustrated in Fig. 3 and shown more explicitly in Appendix A.

$$U(x) = \alpha_1 x^{1/3}, x \geq 0, \alpha_1 > 0 \quad (5)$$

83.5.1 Marginal Utility Is Positive When $x > 0$

Equation (6), the first derivative of Eq. (3), measures a person's marginal utility under favorable circumstances. This derivative is not defined for $x=0$.

$$U'(x) = \frac{d}{dx} [\alpha_1 x^{1/3}] = \frac{\alpha_1}{3x^{2/3}} > 0, \alpha_1 > 0, x > 0 \quad (6)$$

The sign of this first derivative with respect to x is positive, which means that between zero and positive infinity the decision-maker prefers large gains over small gains.

Equation (7) is the second derivative of Eq. (3).

$$U''(x) = \frac{d^2}{dx^2} [\alpha_1 x^{1/3}] = -\frac{2\alpha_1}{9x^{5/3}} < 0, \alpha_1 > 0, x > 0 \quad (7)$$

The negative sign on the second derivative indicates that the cube root utility function has diminishing marginal utility over the range of favorable outcomes. This implies that the cube root utility function in Eq. (3) is concave in the positive quadrant. These equations with these signs represent wealth maximizing risk-averse behavior over the concave line segment.

Equation (8) is the third derivative of Eq. (3), and it is positive.¹⁹

$$U'''(x) = \frac{d^3}{dx^3} [\alpha_1 x^{1/3}] = \frac{10\alpha_1}{27x^{8/3}} > 0, \alpha_1 > 0, x > 0 \quad (8)$$

The derivatives above are used to analyze Absolute Risk-Aversion (ARA) in the next section.

83.5.2 Analyzing Absolute Risk-Aversion (ARA) for Favorable Outcomes, $x > 0$

The Arrow–Pratt Measure of Absolute Risk-Aversion (ARA), named after Kenneth Arrow (1965) and John W. Pratt (1964), measures how an investor's preference for risk changes in response to changes in the investor's wealth. If an investor wants to put more money in risky assets as his/her wealth increases, he/she is said to be

¹⁹The first derivative is positive in Eq. (4), the second derivative is negative in Eq. (5), and the third derivative is positive in Eq. (6), this suggests the investor prefers positive skewness. Later, a Taylor series expansion is employed to make a stronger statement about the decision-maker's preference for positive skewness.

exhibiting decreasing *ARA* (*DARA*). This is normal *DARA* behavior that is more rational and realistic than constant *ARA* (*CARA*) or increasing *ARA* (*IARA*). Eq. (9) defines $ARA(x)$ for the cube root utility function.

$$ARA(x) = -\frac{U''}{U'} = \frac{2}{3x} > 0, \alpha_1 > 0, x > 0 \quad (9)$$

Because Eq. (10) has a negative sign and x is always positive in this range, the cube root utility function always has *DARA* over the range of positive outcomes.

$$ARA'(x) = \frac{d}{dx}[ARA(x)] = \frac{d}{dx}\left[\frac{2}{3x}\right] = -\frac{2}{3x^2} < 0, \alpha_1 > 0, x > 0 \quad (10)$$

83.5.3 Analyzing Relative Risk-Aversion (RRA) When $x > 0$

Arrow-Pratt's concept of Relative Risk-Aversion (*RRA*) also measures how an investor's preference for risk adjusts in response to changes in the investor's wealth. *RRA* differs from *ARA* as suggested by the following relation : $RRA(x) = xARA(x)$. *RRA* for the cube root utility function is indicated in Eq. (11).

$$RRA(x) = xARA(x) = \frac{2}{3} > 0 \quad (11)$$

If an investor wants to invest a greater *proportion* of his/her wealth in risky assets as his/her wealth increases, he/she is said to be exhibiting decreasing *RRA* (*DRRA*). Constant *RRA* (*CRRA*) is more rational and realistic than *DRRA* or increasing *RRA* (*IRRA*) (I. Friend and M. Blume, 1975). Equation (12) shows that over the range of positive outcomes the cube root utility function exhibits *CRRA*. *CRRA* is one of the qualities the cube root utility functions possess and some alternative utility functions lack.

$$RRA'(x) = \frac{d}{dx}[RRA(x)] = \frac{d}{dx}\left[\frac{2}{3}\right] = 0 \quad (12)$$

83.6 Analysis of the Utility Function with Unfavorable Outcomes, $x < 0$

This section focuses on the lower (losses) segment of Eqs.(1) and (3), which is rewritten as Eq. (13).

$$U(x) = \alpha_2 x^{1/3}, x < 0, \alpha_2 > 0 \quad (13)$$

Figure 3 illustrates a symmetric solid gray curve when $\alpha_2 = 1$. And Fig. 3 illustrates the asymmetric dotted curve when $\alpha_2=3.25$.

83.6.1 Analyzing Marginal Utility When $x < 0$

Equation (14) is the first derivative of Eq. (13); it equals the marginal utility of losses in an unfavorable environment. For any value of x , $x^{2/3}$ is always positive; thus, the first derivative is positive.

$$U'(x) = \frac{d}{dx} [\alpha_2 x^{1/3}] = \frac{\alpha_2}{3x^{2/3}} > 0, x < 0, \alpha_2 > 0 \quad (14)$$

This positive marginal utility indicates that if a typical person had to lose money they would always rather have small losses than large losses. This behavior represents rational economic behavior.

Equation (15) is the second derivative of Eq. (13). The positive sign of the second derivative implies that over the range from negative infinity to zero, the decision-maker experiences increasing marginal utility as x increases.

$$U''(x) = \frac{d^2}{dx^2} [\alpha_2 x^{1/3}] = -\frac{2\alpha_2}{9x^{5/3}} > 0, x < 0, \alpha_2 > 0 \quad (15)$$

The range of x values is strictly negative throughout section. For negative values of x , the value of $x^{5/3}$ will also always be negative. This negates the negative sign of the derivative and makes the derivative positive. When Eq. (15) is positive, the left-hand side of the utility curve in Fig. 2 is convex. The convex line segment of the cube root utility function suggests that risk-taking behavior tends to prevail when losses occur.

Equation (16) is the third derivative of Eq. (13)

$$U'''(x) = \frac{d^3}{dx^3} [\alpha_2 x^{1/3}] = \frac{10\alpha_2}{27x^{8/3}} > 0, x < 0, \alpha_2 > 0 \quad (16)$$

For any value of x , the value of $x^{8/3}$ is always positive; so, the third derivative is positive.

83.6.2 Analyzing Absolute Risk-Aversion (ARA) for $x < 0$

As mentioned above, the Arrow–Pratt Measure of Absolute Risk-Aversion (*ARA*) measures how an individual's preference for risk responds to changes in the person's wealth. Equation (17) defines *ARA* for the cube root utility function under unfavorable outcomes.

$$ARA(x) = -\frac{U''}{U'} = \frac{2}{3x} < 0, x < 0, \alpha_2 > 0 \quad (17)$$

The negative sign on Eq. (18) indicates that the individual, while experiencing losses, $x < 0$, will experience diminishing *ARA* (*DARA*).

$$ARA'(x) = \frac{d}{dx}[ARA(x)] = \frac{d}{dx} \left[\frac{2}{3x} \right] = -\frac{2}{3x^2} < 0 \quad (18)$$

Most economists believe *DARA* is logical behavior because investing a larger proportion of one's money in risky assets seems appropriate as a person's wealth increases.

83.6.3 Analyzing Relative Risk-Aversion (RRA) When $x < 0$

As mentioned above, the Arrow-Pratt measure of Relative Risk-Aversion (*RRA*) differs from *ARA* as follows : $RRA = xARA(x)$. Equation (19) defines *RRA* for the cube root utility function over the range of unfavorable outcomes.

$$RRA(x) = xARA(x) = \frac{2}{3} \quad (19)$$

The first derivative of Eq. (19) equals zero, indicating the cube root utility function has constant *RRA* (*CRRA*) over the range of unfavorable outcomes.

$$RRA'(x) = \frac{d}{dx}[RRA(x)] = \frac{d}{dx} \left[\frac{2}{3} \right] = 0 \quad (20)$$

Equation (20) indicates that an investor who is risk-averse under favorable states will become a risk-taker in unfavorable circumstances. The findings above combine to suggest that the cube root utility function is designed to generate wealth maximizing behavior.

83.7 A Taylor Series Expansion of $\alpha_1 x^{1/3}$

The Taylor series expansion in Eq. (19) is used to analyze fluctuations in utility from gains and losses. The Taylor series expansion is used to show that qualified radical utility functions have a consistent preference for positively skewed outcomes.

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0) \frac{(x - x_0)^2}{2!} + f'''(x_0) \frac{(x - x_0)^3}{3!} + R \quad (21)$$

where *R* is a remainder that is omitted in the approximations that follow.

This analysis focuses on the first three moments. Empirical financial research suggests it is not worthwhile to consider more than an investment's first three statistical moments²⁰ (Arditti 1967) (Kraus and Litzenberger 1976). Empirically,

²⁰The subject of Taylor series approximations of utility functions and its application to expected utility in finance has a long history. For a discussion of these applications, see Garlappi and Skoulakis 2011.

the fourth and higher-order moments often have statistically insignificant explanatory power, have signs that alternate from one sample to the next, or have both problems. So, the fourth and higher-order moments are omitted here. Following the suggestion of Levy and Markowitz (1979), which was derived in a slightly different context, we set the expected value of x to be $x_0 = E(x)$ in Eq. (22).

$$f(x) \cong \alpha_1 E(x)^{\frac{1}{3}} + \frac{1}{3} \alpha_1 E(x)^{-\frac{2}{3}} \cdot (x - E(x)) + \frac{1}{3} \left(\frac{-2}{3} \right) \alpha_1 E(x)^{-\frac{5}{3}} \cdot \frac{(x - E(x))^2}{2!} + \frac{1}{3} \left(\frac{-2}{3} \right) \left(\frac{-5}{3} \right) \alpha_1 E(x)^{-\frac{8}{3}} \cdot (x - E(x))^3 / 3! \quad (22)$$

The expected value of Eq. (22) shows that the prospect's expected utility is positive. It is comforting to find that the expected value is positive because it would be irrational economic behavior to invest in a prospect that did not have a positive expected payoff. For the cube root utility function, Equation (22) is reformulated in terms of $\alpha_1 E(x)^{\frac{1}{3}}$ below.

$$E \left[\alpha_1 (x)^{\frac{1}{3}} \right] \cong \alpha_1 E(x)^{\frac{1}{3}} + \alpha_1 \frac{1}{3} \left(\frac{-2}{3} \right) E(x)^{-\frac{5}{3}} \cdot E \left((x - E(x))^2 \right) / 2! + \frac{1}{3} \left(\frac{-2}{3} \right) \times \left(\frac{-5}{3} \right) \alpha_1 E(x)^{-\frac{8}{3}} \cdot E \left((x - E(x))^3 \right) / 3!$$

The following un-numbered equation is a simplified version of the preceding un-numbered equation.

$$\alpha_1 E \left(x^{\frac{1}{3}} \right) \cong \alpha_1 E(x)^{\frac{1}{3}} - \alpha_1 \frac{2}{9} E(x)^{-\frac{5}{3}} \cdot \sigma^2 / 2! + \frac{10}{27} \alpha_1 E(x)^{-\frac{8}{3}} E \left((x - E(x))^3 \right) / 3!$$

The equation above shows that for the cube root utility function, the first and third moments have a positive impact on expected utility, while the variance has a negative impact on expected utility. This last equation shows that outcomes with positive skewness are always preferred over similar outcomes that are not positively skewed.

83.8 Summary and Conclusions

The mathematics in Eqs.(1) through (20) suggest the qualified radical utility functions and, in particular, the cube root utility function, are good economic theories.

The cube root utility function rationalizes the same economic issues that KT's cumulative prospect theory rationalizes, but the vNM model is based on more widely used theories about utility and classical probability theory. These are the utility theories and probability theories used in several other Nobel Prize winning economic theories and in other social sciences as well.

83.8.1 Summary Points About Cube Root Utility Theory

The infinite number of qualified radical utility functions discussed in this chapter contains the following desirable characteristics.

- (i) The cube root function and the similar radical functions are framed to include both a concave line segment and a convex line segment in a single equation.
- (ii) A countably infinite number (all equations with odd exponents) of other qualified radical utility functions have the same desirable characteristics as the cube root function that was analyzed in Eqs.(1)–(20). A few of the most plausible qualified radical utility functions are analyzed further in Appendix A.
- (iii) The cube root utility function can span a wider range of utility outcomes than the other qualified radical utility functions, as shown in Fig. 5. For this reason, the cube root utility function is considered more desirable than the other qualified radical utility functions.
- (iv) Using vNM utility with probabilities that sum to one enables us to use the mathematical expectation operator, standard deviation, variance-covariance matrix, and other important classical statistical tools that cannot be used in KT's 1979 and 1992 prospect theories.
- (v) The general formula Eq. (3) can be reformulated to assume varying shapes to model an infinite number of different of economic behaviors.

83.8.2 The Economics

The large family of qualified radical utility functions each generate the following kinds of desirable economic behavior:

- (i) Positive marginal utility of wealth and returns over the real set.
- (ii) Consistent risk-aversion over the concave (positive) segment of the real set.
- (iii) Consistent risk seeking over the convex (negative) segment of the real set.
- (iv) Consistent preference for positively skewed outcomes.
- (v) Rationalization of loss aversion when it is formulated to be appropriately asymmetric.
- (vi) All other complicated economic behaviors from Kahneman and Tversky's cumulative prospect theory (Tversky and Kahneman 1992), with the exception of the weighted probabilities.

83.8.3 Comparing Utility Functions

Table 1 summarizes the good and bad characteristics of some popular utility functions, which compete with the cube root function.

All of the criterion listed in column one of Table 1 have the most favorable results for the cube root function, which suggests it is more useful for utility analysis than the alternative utility functions.

Showing that the qualified radical utility functions analyzed above generate behaviors that align with a wide range of desirable outcomes suggests that the vNM theory is more likely to find useful applications in economics, finance, decision theory, and psychology than the competing utility functions. The qualified radical utility functions are also more amenable to various adaptations and linkages than KT's cumulative prospect theory because the objective vNM probabilities sum to one. This chapter concludes that it would be advantageous for behavioral economics research with weighted probabilities to be carried out separately from research involving vNM models. Separating these two important behavioral economic models would not detract from either one, but it could make each one more useful.

Table 1 Summary of advantages and deficiencies of popular utility functions

Criteria for a useful utility function	Five competing utility functions				
	Cube root utility function	Quadratic utility function	Square root utility function	Log utility function	Exponential utility function
<i>Does x span the real line between $-\infty$ to $+\infty$?</i>	Yes, with a single equation	Yes, if it faces down	No, it is undefined for losses	No, it is undefined for losses	No, it is undefined for losses
<i>A positive monotone function, $\frac{dU(x)}{dx} > 0$?</i>	Yes, more wealth is always preferable	Yes, but in only one quadrant	No, it is undefined for losses	No, it is undefined for losses	No, it is undefined for losses
<i>A consistent preference for positive skewness?</i>	Yes	Not consistently	No skewness preference	No skewness preference	No skewness preference
<i>Can losses, $x < 0$, be analyzed?</i>	Yes	Yes	Not able to analyze losses	Not able to analyze losses	Not able to analyze losses
<i>Absolute risk aversion (ARA)</i>	DARA is desired	IARA is not desirable	DARA is desired	DARA is desired	CARA is not desirable
<i>Relative risk aversion (RRA)</i>	CRRA is desired	CRRA is desired	CRRA is desired	CRRA is desired	IRRA is not desirable
<i>Always represents rational economic behavior?</i>	Yes, it is never irrational	Not all segments behave rationally	No, undefined for losses	No, undefined for losses	No, CARA always exists
<i>Can this vNM function explain the complicated behaviors of KT's prospect theory?</i>	Yes, except the weighted probabilities	No, it has many shortcomings	No, it has many shortcomings	No, it has many shortcomings	No, it has many shortcomings
<i>Total desirable yes answers</i>	8 yes answers	3 yes answers	2 yes answers	2 yes answers	Zero yes answers

Appendix A – Further analysis of radical functions

Table A shows the first three derivatives of the first three radical utility functions. These derivatives provide necessary and sufficient conditions to align with similar findings from KT’s cumulative prospect theory.

The numerous qualified radical utility functions with an inflection point at the origin that are discussed above can replicate the complicated economic behaviors generated by KT’s cumulative prospect theory. These qualified radical utility functions differ from KT’s cumulative prospect theory only because they are all based on the vNM probability assumptions instead of including KT’s weighted probabilities.

Appendix Table A – Derivatives of qualified radical utility functions

	$U(x) = \sqrt[3]{x} = x^{1/3}$	$U(x) = \sqrt[5]{x} = x^{1/5}$	$U(x) = \sqrt[7]{x} = x^{1/7}$
1) dU/dx	$\frac{1}{3x^{2/3}}$	$\frac{1}{5x^{4/5}}$	$\frac{1}{7x^{6/7}}$
2) d^2U/dx^2	$-\frac{2}{9x^{5/3}}$	$-\frac{4}{25x^{9/5}}$	$-\frac{6}{49x^{13/7}}$
3) d^3U/dx^3	$\frac{10}{27x^{8/3}}$	$\frac{36}{125x^{14/5}}$	$\frac{78}{343x^{20/7}}$

Notes: The first two derivatives of radical utility functions with an odd (but not even) number in the denominator of the exponent have diminishing positive marginal utility. In addition, the third derivatives combined with the Taylor series reveals that the decision-maker always prefers positively skewed outcomes. Table A is not extended to include an infinite number of additional columns, but it is possible to get similar results for exponents of 1/9, 1/11, 1/13, 1/15, ... ad infinitum.

References

Arditti, Fred D. 1967. Risk and the Required Return on Equity. *Journal of Finance*.
Arrow, Kenneth J. 1965. Aspects of the Theory of Risk Bearing. *The Theory of Risk Aversion*. Helsinki: Yrjö Jahnssonin Saatio. Reprinted in: *Essays in the Theory of Risk Bearing*, Markham Publication Co., Chicago, 1971, 90–109.
Barberis, Nicholas, and Ming Huang. 2001a. Mental Accounting, Loss Aversion, and Individual Stock Returns. *Journal of Finance* 56 (4): 1247–1292.
Barberis, Nicholas and Ming Huang. 2001b. “Mental Accounting, Loss Aversion, and Individual Stock Returns.” *Journal of Finance*, 56(4), see 1251–1259.
Barberis, Nicholas, and Wei Xiong. 2012. Realization Utility. *Journal of Financial Economics* 104 (2): 251–271.
Barberis, Nicholas, Ming Huang, and Richard H. Thaler. 2006. Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing. *American Economic Review*. 96 (4): 1069–1090.
Barberis, Nicholas, Abhiroop Mukherjee, and Baolian Wang. 2016. Prospect Theory and Stock Returns: An Empirical Test. *Review of Financial Studies*. February Issue.
Barberis, Nicholas, Lawrence Jin, and Baolian Wang. 2020. Prospect Theory and Stock Market Anomalies, NBER Working Paper 27155, DOI <https://doi.org/10.3386/w27155>.

- Benartzi, Shlomo, and Richard H. Thaler. 1995. Myopic Loss Aversion and the Equity Premium Puzzle. *Quarterly Journal of Economics*. 110 (1): 73–92.
- Birru, Justin, and Baolian Wang. 2016. Nominal Price Illusion. *Journal of Financial Economics* 119 (3): 578–598.
- Black, Fischer, and Myron Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*. 81 (3): 637–654.
- Conrad, Jennifer, Nishad Kapadia, and Yuhang Xing. 2014a. Why Do Individual Investors Hold Overpriced Stocks? *Journal of Financial Economics*: 455–475.
- . 2014b. Death and jackpot: Why do individual investors hold overpriced stocks? *Journal of Financial Economics* 113 (3): 455–475.
- Francis, Jack Clark. 2021. Reformulating Prospect Theory to Become a von Neumann-Morgenstern Theory. *Review of Quantitative Finance and Accounting* 56 (1): 965–985.
- Frazzini, Andrea. 2006. The Disposition Effect and Underreaction to News. *Journal of Finance*. 61 (4): 2017–2046.
- Friend, Irwin, and Marshall E. Blume. 1975. The Demand for Risky Assets. *American Economic Review* 65 (5): 900–922.
- Frydman, Cary, Samuel M. Hartzmark, and David H. Solomon. 2018a. Rolling Mental Accounts. *The Review of Financial Studies* 31 (1): 362–397.
- Frydman, Cary, Samuel M. Hartzmark, and David Solomon. 2018b. Rolling Mental Accounting. *Review of Financial Studies*., January Issue 31: 263–291.
- Garlappi, Lorenzo, and Georgios Skoulakis. 2011. A State Variable Decomposition Approach for Solving Portfolio Choice Problems. *Mathematics and Financial Economics*. 5 (2): 121–156.
- Grinblatt, M., and B. Han. 2005. Prospect Theory, Mental Accounting, and Momentum. *Journal of Financial Economics* 78: 311–339.
- Harel, Arie, J.C. Francis, and Giora Harpaz. 2018. Alternative Utility Functions: Review, Analysis and Comparison. *Review of Quantitative Finance and Accounting*., October Issue 51: 785–811.
- Imas, Alex. 2016. The Realization Effect: Risk-Taking after Realized versus Paper Losses. *American Economic Review* 106 (8): 2086–2109.
- Justin, Birru, and Baolian Wang. 2016. Nominal Price Illusion. *Journal of Financial Economics* 119 (3): 578–598.
- Kahneman, Daniel, and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*. 47 (2): 263–291.
- . 2000. *Choices, Values and Frames*. New York: Cambridge University Press.
- Kraus, A., and R. Litzenberger. 1976. Skewness Preference and the Valuation of Risky Assets. *Journal of Finance* 21: 1085–1094.
- Levy, H., and Harry M. Markowitz. 1979. Approximating Expected Utility by a Function of Mean and Variance. *American Economic Review* 69 (3) June Issue: 308–317.
- Li, Yan, and Liyan Yang. 2013. Prospect Theory, the Disposition Effect and Asset Prices. *Journal of Financial Economics*., March Issue 107 (3).
- Markowitz, Harry. 1952. The Utility of Wealth. *Journal of Political Economy* 60 (2): 151–158.
- Markowitz, H.M. 1952b. Portfolio Selection. *The Journal of Finance*. 7 (1): 77–91. <https://doi.org/10.2307/2975974>. JSTOR2975974.
- Mehra, Rajnish, and Edward C. Prescott. 1985. The Equity Premium: A Puzzle. *Journal of Monetary Economics* 15 (2): 145–161.
- Odean, Terrance. 1998a. Volume, Volatility, Price, and Profit When All Traders Are Above Average. *Journal of Finance* LIII (6) December Issue: 1887–1934.
- . 1998b. Are Investors Reluctant to Realize Their Losses? *Journal of Finance*. 53 (5): 1775–1798.
- Pratt, J.W. 1964. Risk Aversion in the Small and in the Large. *Econometrica*. 32 (1–2): 122–136.
- Sharpe, William F. 1964. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance*. 19 (3): 425–442.

- Shefrin, Hersh, and Meir Statman. 1985. The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *Journal of Finance.*, July Issue 40 (3): 777–790. <https://doi.org/10.1111/j.1540-6261.1985.tb05002.x>.
- Statman, Meir, Steven Thorley, and Keith Vorkink. 2006. Investor Overconfidence and Trading Volume. *The Review of Financial Studies* 19 (4) Winter Issue: 1531–1565. <https://doi.org/10.1093/rfs/hhj032>.
- Thaler, Richard H. 1985. Mental Accounting and Consumer Choice. *Marketing Science* 4 (3): 199–214.
- Tversky, Amos, and Daniel Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211 (4481): 453–458. <https://doi.org/10.1126/science.7455683>.
- . 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5 (4): 297–323.
- von Neumann, Jon, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press. Second ed. 1947, Third ed. 1953.



A Global Comparative Study of Impact Investments Research in Academic Institutions

84

Rachel Calipha and Itzhak Venezia

Abstract

This chapter analyzes the comparative contributions of academia to the development of social impact investment in recent years. This subject is quite fashionable and relatively new, so it is important to explore the ways and directions in which it is being promoted by the academy. To this end, we investigated the following issues: how many papers were written about social impact investment in each of the last several years (as a metric of academic evolution), the relative involvement of public versus private sector universities (presumably the former should be more concerned about social issues), which academic disciplines are active in this area (as an indication of the backgrounds and perspectives of the researchers in the field), the distribution of the methods of analysis (mainly to explore to what extent the knowledge is grounded in theory and empirical results), and the extent to which the academic papers were funded by external sources (as a crude measure of the authors' independence). We also documented the papers published in this area and catalogued them according to their various disciplines. A list of sponsors who contributed to these papers is provided as a database for researchers who are seeking financial aid. Finally, the academic influence of these papers is examined.

R. Calipha

School of Management and Economics, The Academic College of Tel Aviv-Yaffo, Tel Aviv, Israel
e-mail: Rachelca@mta.ac.il

I. Venezia (✉)

School of Business, The Hebrew University, Jerusalem, Israel

The Hebrew University of Jerusalem, Jerusalem, Israel

The Academic College of Tel Aviv-Yaffo, Tel Aviv, Israel

e-mail: itzhakve@mta.ac.il

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_85

1963

Keywords

Impact Investing · Social Impact Investment · Social Impact Bonds. Public Universities vs. Private Universities · Funding · h-index

JEL Classification

A13 · G11 · M14

What's in a name? That which we call a rose, by any other name would smell as sweet.
William Shakespeare, *Romeo and Juliet*

84.1 Introduction

The field of impact investing has grown significantly in recent years. The annual impact investor survey conducted by the Global Impact Investment Network (GIIN) shows that the value of the assets managed by impact investing organizations has grown from \$77.4 billion in 2015 (GIIN 2016) to \$404 billion at the end of 2019 (GIIN 2020). In addition, the Global Steering Group (GSG), a group that catalyzes impact investment, went from representing 15 countries and the EU in 2015 (EVPA 2017) to representing 33 countries and the EU in 2020 (GSG 2021). In the same vein, the 2020 World Investment Report (UNCTAD 2020, p. 187) stated that the value of “sustainability-dedicated investments” (which include impact investments) is estimated in the range of \$1.2 to \$1.3 trillion at the end of 2020.

The speed at which this area is evolving calls for a frequent review of the ways and directions in which it is advancing. Therefore, in this chapter we examine the pace and direction of the dissemination of new academic knowledge in impact investing in the hope that this knowledge will contribute to the assessment of its future growth and trajectory. A previous comprehensive review of the scope of research in this field was conducted by Daggers and Nicholls (2016), who examined the period between 2010 and September 2015.¹ The current study adds to the literature in the following ways. First, in light of the rapid pace at which the field is expanding (according to the GIIN 2020 data above, impact investments have grown at least five-fold since the previous study), a frequent review of it is recommended. Second, we sharpen the previous knowledge by studying more detailed data. Doing so enables us to explore several issues not hitherto examined, e.g., the relative involvement of public versus private sector universities in this field; the academic disciplines active in this area; the distribution of the methods of analysis; and the extent to which the academic papers were funded by external

¹Daggers and Nicholls's (2016) paper followed other review studies (see, e.g., Nicholls and Pharoah 2008; Lyons and Kickul 2013; Höchstädter and Scheck 2015; Clarkin and Cangioni 2016; Rizzello et al. 2016; and Friedrich-Silber and Rosenhek 2000). Broccardo and Mazzuca 2019, provided a review but studied only social impact bonds and Kubátová and Kročil 2020 did not use the umbrella term SII.

sources. Third, we provide a list of the papers we found in this area arranged by their various disciplines. Fourth, a list of sponsors who contributed to these papers is provided as a database that will serve researchers who seek financial aid. Finally, we examine the academic impact of the research in this area.

The chapter is constructed as follows. In Sect. 84.2, we present the terminology of impact investments and related activities. In Sect. 84.3, we present our methodology and data sources. In Sect. 84.4, the results are presented, and in Sect. 84.5 we present our conclusions

84.2 The Terminology of Impact Investments and Related Activities

A representative recent definition of impact investments was provided by the GIIN (2020): “Impact investments are investments made with the intention to generate positive, measurable social and environmental impact alongside a financial return.” This term follows other similar terms that represent analogous concepts, such as sustainability, socially responsible investment (SRI), corporate social responsibility (CSR), social investment (SI), environmental, social and governance (ESG), social impact bonds (SIB) and social impact investment (SII). All of these terms refer to activities aimed at incorporating externalities into the decisions of individuals and firms to optimize the welfare of society and not only that of the individual. The terms may differ in their aims, means, scopes, and measures but are quite closely related, so it is difficult to separate one from the other. Moreover, new, related terms are continuously being proposed. For example, the term “integral investing” (Kubátová and Kročil 2020), which adds aims and metrics to impact investing, recently emerged.²

This study focuses on Social impact investment (SII). SII (see, e.g., Daggers and Nicholls [2016]) is an umbrella term that encompasses the terms *impact investing* and *social investment*. It is important to distinguish between these two terms. The first was proposed by the Rockefeller Foundation in 2007 in a strategic attempt to create a momentum of positive investment with social and environmental impact in the USA. The second, in contrast, was developed earlier in a less strategic environment, mainly in Europe. Another distinction between the terms is their focus (Daggers and Nicholls 2016, p. 6). Impact investing involves the use of capital to create specific social or environmental impact, whether through direct allocation of capital, investment in funds, or contractual agreements such as SIBs. Its focus is therefore mainly on investor behavior and motivations. Social investment involves providing social sector organizations (SSO) access to repayable

²This term adds positive cultural and behavioral aspects and human development to the basic aims of impact investment (which are only profit and positive societal and/or environmental impact). Integral investing also suggests adding team culture, ethics, morals, and personal values of team members to the basic metrics of profits and ESG criteria required by impact investing.

capital whose providers are motivated to create social or environmental impact. As a result, there is more of focus on the investee.

84.3 Data and Methodology

The study was conducted by means of two main databases that list papers published in various areas: SCOPUS and Google Scholar. Papers that address social impact investment were chosen. The period reviewed was from 2015 to December 2020. A search for the keywords *impact investing* (II), *social investment* (SI), *social impact bonds* (SIB), and *social impact investment* (SII) in abstracts or titles was conducted. The final sample consists of 119 papers published in journals or working papers, but excludes books. Classification into the various categories (discipline, method of analysis, academic affiliations of the authors, and whether the research received external funding) was conducted subjectively by the authors based on the titles, abstracts, and acknowledgments sections of the chosen papers. When these sections did not provide sufficient information, the entire paper was read. The classification of universities in the USA into private or public was inferred from the website UnivSearch: US College and University Directory (<http://www.univsearch.com/state.php>) and the classification of European universities was taken from Wikipedia (https://en.wikipedia.org/wiki/Category:Private_universities_and_colleges_in_Europe). Finally, to examine the academic influence of research on social impact investment research we measured the h-index. This index expresses the number of articles (h) in a journal that have received at least h citations. It quantifies both journals' scientific productivity and their scientific impact and is also applicable to scientists, countries, and so on. The index was taken from Scimago Journal & Country Rank (<https://www.scimagojr.com/>).

84.4 Results and Discussion

In what follows, we present the findings of our study in six subsections, each of which corresponds to one of the different topics we investigated.

84.4.1 Evolution of the Field Over Time

Table 1, below, shows the yearly number of papers published during our study period. We cannot detect a statistically significant trend from this table, possibly due to the small number of years studied, although a visual inspection reveals a slightly positive trend. This modest increase in publications stands in contrast to the rapid growth in impact investments mentioned above and to the more detailed data published by the GIIN (2020) shown in Table 1.

Table 1 Number of papers and volume of Assets under Management by year

Year	Number of papers	Assets under Management (AUM) ^b
2020 ^a	30	N/A
2019	32	404
2018	23	239
2017	10	228.1
2016	8	114
2015	16	77.4

^aPartial year, until November 2020

^bIn billions of dollars. Source: GIIN (2020). AUM refers to the total market value of the investments or assets managed by fund managers, DFIs (government-backed entities), bank/diversified financial institutions, foundations, family offices, pension funds, insurance companies, and others.

84.4.2 Distribution of the Research Institutions (Public University, Private University, or Nonacademic Institution)

The country where the research was conducted can give us some indication of which countries are most concerned with the subject of social impact investment. Scientifically and economically developed countries would be expected to produce more research on impact investment than less developed ones. In this regard, the only surprising finding is that the USA, the country that leads the world in research publications in most areas, lags behind the UK in terms of research on social impact investment. The finding that research of the subject is concentrated in only two countries that together conduct more research than all the others combined is also striking. Our findings are consistent with those of Daggers and Nicholls (2016), with the exception of the growth of publications in Italy in recent years.

In addition, most of the publications were written by authors from public universities, except for those published in the USA and Germany. Whereas in the USA the number of private universities (for-profit and nonprofit) greatly exceeds that of public ones (2672 vs. 1626 in 2019, or 62% vs. 38%), they are almost evenly matched in terms of research on impact investments. But because public universities are usually much larger than private ones (enrollment in public universities in the USA was 14.9 million students in 2019 vs. only 5.2 in private universities, [Statista 2020]), we cannot conclude which type of institution is more concerned with impact investments (Table 2).

84.4.3 Relative Weight of the Academic Disciplines Involved in Impact Investment Research

In Table 3, we present the distribution of papers among the various academic disciplines involved in the study of social impact investment. By definition, impact investments are concerned with investing and hence with finance, economics, business, and management. Since impact investments also concern externalities

Table 2 Number of universities by country and type (private vs. public) and nonacademic institutions

Country	Number of universities ^a	Private	Public	Nonacademic institutions
England	40	0	40	1
USA	21	11	10	1
Italy	14	2	12	0
Canada	5	0	5	0
Germany	5	3	2	0
Australia	4	0	4	0
France	4	1	3	0
Switzerland	4	0	4	0
Austria	3	0	3	0
South Africa	3	0	3	1
Korea	2	0	2	0
Russia	2	0	2	0
Netherlands	2	0	2	0
Czech Republic	1	0	1	0
Denmark	1	0	1	0
Hong Kong	1	0	1	0
Kazakhstan	1	0	1	0
Luxembourg	1	0	1	0
Malaysia	1	0	1	0
Norway	1	0	1	0
Spain	1	1	0	1
Turkey	0	0	0	1
Total	114	18	96	5

^aThe university is determined by the main author’s affiliation.

Table 3 Distribution of social impact investment papers among the various disciplines

	Finance and Economics	Business and Management	Public Policy	Social Policy
Number of articles	23	52	14	24
Percentage	20.4%	46.0%	12.4%	21.2%

and society, they are related to public finance and policy. It thus comes as no surprise that most papers are published in journals associated with these disciplines. The list of papers categorized by discipline is presented in [Appendix 2](#). This list could be useful to researchers who seek to focus on a particular discipline. It could also serve as a benchmark for future studies that examine the direction that impact investment has taken.

84.4.3.1 Types of Analysis Used in Social Impact Investment Research

In [Table 4](#), we present the distribution of papers according to the types of analysis they employ: theoretical, scoping, qualitative, and empirical. An interesting insight

Table 4 Allocation of social impact investment papers by method of analysis

	Theoretical	Scoping	Qualitative	Empirical
Number of articles	19	31	56	13
Percentage	16.0%	26.1%	47.1%	10.9%

Table 5 Number of funders by funding origin

Funding origin	Number of funders
UK	10
European Union	6
Canada	6
USA	4
Italy	4
Korea	3
South Africa	2
Abu Dhabi	1
Czech Republic	1
Spain	1
Belgium	1
Total	39 ^a

^aThe number of funders is greater than the number of papers since four of the papers were sponsored by more than one funder

that emerges from this table is the paucity of empirical research (11%) relative to scoping (26%) and qualitative analysis (47%). One reason for this imbalance is that the field is still in its infancy and there remains a need to delineate its objectives and the possible means of reaching them, to clarify which organizations should be examined, to identify the parties that may gain or lose from impact investments, and so on. Since, theoretically, impact investment has unquantifiable pros and cons, its net effects on welfare are only a matter of empirical findings (i.e., real data pertaining to the theories and assertions regarding impact investing; see, e.g., Spiess-Knafl and Aschari-Lincoln 2015; Karpenko and Voronzhak 2018; Mersland et al. 2020). It therefore may seem that additional empirical research is needed. Notably, however, since the period of the previous survey by Daggers and Nicholls (2016), several data bases have been constructed and are available for research, thus increasing the likelihood that the relative intensity of empirical research will grow in the future.

84.4.4 Funding

Table 5, below, which lists the number of funded papers in all countries, shows that only 29% (35/119) of the papers were funded, and the sources of funding were mainly from the UK, Canada, and the European Union (other than the UK). Most of the funds came from governments (see Appendix 1 for details), a fact that indicates their higher relative concern with social impact investments.

Table 6 Number of papers by journal and h-index

Journal ^a	h-Index	Number of papers
<i>Public Money & Management</i>	46	9
<i>Journal of Urban Affairs</i>	57	5
<i>Journal of Social Entrepreneurship</i>	23	5
<i>Journal of Business Ethics</i>	168	3
<i>Research in International Business and Finance</i>	35	3
<i>Journal of Comparative Policy Analysis: Research and Practice</i>	21	3
<i>Journal of Economic Policy Reform</i>	20	3
<i>Nonprofit and Voluntary Sector Quarterly</i>	76	2
<i>Social Policy & Administration</i>	61	2
<i>Policy & Politics</i>	50	2
<i>Education Policy Analysis Archives</i>	44	2
<i>Policy and Society</i>	31	2
<i>Social Responsibility Journal</i>	27	2
<i>Journal of Applied Accounting</i>	19	2
<i>Journal of Business Venturing Insights</i>	14	2
<i>Journal of Sustainable Finance & Investment</i>	13	2
<i>Entrepreneurship Research Journal</i>	11	2
<i>Economy and Space</i>	0	2
<i>International Journal of Voluntary and Nonprofit Organizations</i>	0	2

^aJournals with fewer than two publications are not presented.

84.4.5 Academic Influence of Social Impact Investment Research

The findings presented in Table 6 show that articles that address social impact investments have been published in a variety of journals and there is no one main journal associated with the field. Another notable observation is that papers on social impact investments are published in journals with low h-indices (only three papers were published in journals with an h-index higher than 100). This may be due to fact that the field is in its early stages of development and/or to the fact that it comprises a very specific research niche area.

84.5 Conclusions

From our findings, we observe that the growth of impact investments over the last few years has been much more rapid than the academic developments in this area. Moreover, empirical studies on the subject comprise a paltry 11% of total research. Nonetheless, this type of research is crucial if one is to convince others of the benefits or disadvantages of impact investing. Thus, the influence of academia on social impact investment seems to hinge to a great extent on its ability to produce meaningful empirical research in this area. This, in turn, requires the existence of extensive databases, the lack of which could be partly to blame for the current paucity of empirical studies (see, e.g., Reisman et al. 2018; Jamieson et al. 2020).

Furthermore, the dearth of empirical research could be part of the reason for the weak academic influence exhibited by the relatively low h-indices of the journals where papers in the area are published. Most leading journals in economics and finance insist that authors publishing in their journals support their claims with sound empirical analysis. It would be difficult for the area to become mainstream without the respectability gained by publishing papers in leading journals, and in order for this to happen, databases are needed to facilitate the kind of analysis that these journals demand.

In addition, our findings indicate that the UK and the USA are dominant in this area in all respects: the number of papers published, the number of universities involved with these publications, and the sponsorship (especially the UK). This is hardly surprising since these are developed countries with mature SII markets and well-established SII public policies (OECD 2019). What is somewhat surprising is the relatively meager participation of other developed countries in publications and the absence of developing countries, who have the most to gain from SII.

Universities and governments both play important roles in economic development. Universities' roles include teaching, conducting research, and serving as innovative engines for economic development (Geiger 1986, 1993; Leslie and Caldwell 1990; Leslie 1993; Florida et al. 2006). The role of governments is to set policies and, where necessary, provide funds to help to achieve its goals (Stiglitz 1997; Adelman 2000). An important conundrum is whether research (universities) or policy (government) comes first and what the optimal division of tasks between them is.

In the social sciences (as opposed to the natural sciences and medicine), practice usually precedes research, so that it is common to engage in studying and conceptualizing a new social phenomenon after it has been taking place in the real world, accompanied by policy (Calipha and Gidron 2020). The evolution of the third sector usually follows a different path, with research and conceptualization preceding practice and policy. This is true for many countries, but in the case of the Israeli third sector, as we show below, this pattern is particularly striking. One wonders how the fact that impact investments are business/nonprofit hybrids (i.e., their agendas include social goals in addition to profits) will influence the patterns of their evolution. Observing the development of the third sector in Israel may provide insights regarding this question.

In Israel, research activity on the third sector began in the 1970s and 1980s, when foreign and local researchers began to study related subjects.³ Then, in the 1990s, came international conferences that led to joint international projects that provided tools for defining the field and facilitated international comparisons.⁴ The projects' outcomes accelerated further academic research, and so on.⁵ It is worth noting that during this period, NGOs contributed significantly to society and their number

³See, e.g., Kramer 1970, 1976, 1981, 1984; Katan 1988; Eran 1992; Jaffe 1979, 1982, 1983, 1985, 1987; Yishai 1987; Gidron 1984; Gidron and Bargal 1986; Gidron and Chesler 1994.

⁴Salamon et al. (1999).

⁵See, e.g., Bar-Mor 1999; Gidron 1999; Gidron and Katz 1998; Gidron et al. 2003; Shay et al. 1999; Silber and Rosenhek, 2000; Telias et al. 2000.

significantly increased over the years. The influence of academia became more significant only later in the process, after academic researchers managed to develop local databases and enhanced government involvement through joint meetings that eventually paved the way for the field to enter the mainstream.⁶ Today, the third sector in Israel has an impressive presence in society (Calipha and Gidron 2020).

Only time will tell whether social impact investments will develop in a similar manner, but in any case, it seems that the strong involvement of academia in this area is crucial for its advancement. We hope this chapter will assist future researchers to propose ways in which academia can work to increase its effect on social impact investments.

Appendix 1: List of Sponsors

Country	Funding organizations	Type of funders
UK		
1	Economic and Social Research Council in the United Kingdom	Government
2	The University of Warwick	University
3	NIHR Policy Research Programme*	Government
4	UK government, including the Cabinet Office and Department of Health	Government
5	Policy Research Programme of the Department of Health for England*	Government
6	The Government Outcomes Lab (GO Lab)	University
7	The Welsh Government through Health and Care Research Wales	Government
8	A partnership between the UK Department for Digital, Culture, Media and Sport and the Blavatnik School of Government at the University of Oxford	Government
European Union		
9	The BMW Foundation Herbert Quandt	Foundation
10	European Investment Bank	Government
11	The EU FP7 project “Creating Economic Space for Social Innovation”	Government
12	Luigi Einaudi Chair Innovation Grant from the Cornell Institute for European Studies	University
13	European Commission	Government
14	The European Union’s EIT Climate KIC	Government
Canada		
15	Macquarie Group Foundation	Foundation
16	Social Science and Humanities Research Council of Canada ^b	Government
17	The University of Sydney	University

(continued)

⁶See, e.g., Galnoor 2003.

Country	Funding organizations	Type of funders
18	CMA Ontario	University
USA		
19	United States Department of Agriculture (USDA)	Government
20	Helene and Grant Wilson Center for Social Entrepreneurship, Pace University	University
21	Atkinson Center for Sustainable Futures	University
22	US Department of Education (Grant No. U411B110098) and the National Institute of Child Health and Human Development	Government
Italy		
23	Project Social Impact Finance (SIF16_00055) “An Italian Platform for Impact Finance: Financial Models for Social Inclusion and Sustainable Welfare”(in collaboration with Sapienza)	University
24	Italian Ministry of Education, Universities and Research (MIUR) ^a	Government
25	Project Social Impact Finance (SIF16_00055) “An Italian Platform for Impact Finance: Financial Models for Social Inclusion and Sustainable Welfare” (in collaboration with Sapienza)- Italian Ministry of the Education, University and Research	Government
Korea		
26	National Research Foundation of Korea ^a	Government
27	The Korea Foundation	Foundation
South Africa		
28	University of Johannesburg	University
29	Global Fund to Fight AIDS, TB and Malaria: The South African Department of Science and Technology	Government
Abu Dhabi		
30	Abu Dhabi Education Council, INSEAD Randomized Control Trials Lab	Government
Czech Republic		
31	Ministry of Education, Youth and Sports, Czech Republic	Government
Spain		
32	ROCHE FARMA, SA	Private
Belgium		
33	Skopos Impact Fund	Foundation

Appendix 2: Impact Investment Literature Catalogued by Discipline

Finance and Economics

Alijani, S., & Karyotis, C. (2019). Coping with impact investing antagonistic objectives: A multistakeholder approach. *Research in International Business and Finance*, 47, 10–17.

Berndt, C., & Wirth, M. (2019). Struggling for the moral market: Economic knowledge, diverse markets, and market borders. *Economic Geography*, 95(3), 288–309.

(continued)

-
- Broccardo, E., & Mazzuca, M. (2019). The missing link? Finance, public services, and co-production: The case of social impact bonds (SIBs). *Public Money & Management*, 39(4), 262–270.
-
- Carè, R., Rania, F., & De Lisa, R. (2020). Critical success factors, motivations, and risks in social impact bonds. *Sustainability*, 12(18), 7291. <https://doi.org/10.3390/su12187291>
-
- Chowdhry, B., Davies, S. W., & Waters, B. (2019). Investing for impact. *The Review of Financial Studies*, 32(3), 864–904.
-
- Del Giudice, A., & Migliavacca, M. (2019). Social impact bonds and institutional investors: An empirical analysis of a complicated relationship. *Nonprofit and Voluntary Sector Quarterly*, 48(1), 50–70.
-
- Fraser, A., Tan, S., & Mays, N. (2019). To SIB or not to SIB? A comparative analysis of the commissioning processes of two proposed health-focused social impact bond financed interventions in England. *Journal of Economic Policy Reform*, 1–16. <https://doi.org/10.1080/17487870.2019.1572508>
-
- Giacomantonio, C. (2017). Grant-maximizing but not money-making: A simple decision-tree analysis for social impact bonds. *Journal of Social Entrepreneurship*, 8(1), 47–66.
-
- Gray, J., Ashburn, N., Douglas, H., Jeffers, J., Musto, D. K., & Geczy, C. (2015). Great expectations: Mission preservation and financial performance in impact investing. *Available at SSRN 2694620*.
-
- Jafri, J. (2019). When billions meet trillions: Impact investing and shadow banking in Pakistan. *Review of International Political Economy*, 26(3), 520–544.
-
- Kölbel, J. F., Heeb, F., Paetzold, F., & Busch, T. (2020). Can sustainable investing save the world? Reviewing the mechanisms of investor impact. *Organization & Environment*, 33(4), 554–574.
-
- Koshovets, O., & Frolov, I. (2015). Impact investing as a ‘basic innovation’ for the global economy and finance system post-crisis transformation. *Economy & Business*, 9.
-
- Kubátová, J., & Kročil, O. (2020). The potential of impact and integral investing for sustainable social development and the role of academia in their dissemination. *Sustainability*, 12(17), 6939. <https://doi.org/10.3390/su12176939>
-
- Lee, M., Adbi, A., & Singh, J. (2020). Categorical cognition and outcome efficiency in impact investing decisions. *Strategic Management Journal*, 41(1), 86–107.
-
- Mersland, R., Nyarko, S. A., & Sirisena, A. B. (2020). A hybrid approach to international market selection: The case of impact investing organizations. *International Business Review*, 29(1), 101624. DOI: 10.1016/j.ibusrev.2019.101624
-
- Murzaliyeva, A., Markhayeva, B., Bakhyt, E., & Karshalova, A. (2018). Social return on investment as instrument for assessing socially responsible investing in the Republic of Kazakhstan. *Journal of Applied Economic Sciences*, 13(5), 1357–1365.
-
- Nazari Chamaki, F., Jenkins, G. P., & Hashemi, M. (2019). Social impact bonds: Implementation, evaluation, and monitoring. *International Journal of Public Administration*, 42(4), 289–297.
-
- Reeder, N., Colantonio, A., Loder, J., & Rocyn Jones, G. (2015). Measuring impact in impact investing: An analysis of the predominant strength that is also its greatest weakness. *Journal of Sustainable Finance & Investment*, 5(3), 136–154.
-
- Rizzi, F., Pellegrini, C., & Battaglia, M. (2018). The structuring of social finance: Emerging approaches for supporting environmentally and socially impactful projects. *Journal of Cleaner Production*, 170, 805–817.
-
- Scognamiglio, E., Di Lorenzo, E., Sibillo, M., & Trotta, A. (2019). Social uncertainty evaluation in social impact bonds: Review and framework. *Research in International Business and Finance*, 47, 40–56.
-

(continued)

Snider, A. (2015). Impact investing: The performance realities. *Wealth Management Institute*, 1–14.

Spiess-Knafl, W., & Aschari-Lincoln, J. (2015). Understanding mechanisms in the social investment market: What are venture philanthropy funds financing and how? *Journal of Sustainable Finance & Investment*, 5(3), 103–120.

Vecchi, V., & Casalini, F. (2019). Is a social empowerment of PPP for infrastructure delivery possible? Lessons from social impact bonds. *Annals of Public and Cooperative Economics*, 90(2), 353–369.

Business and Management

Agrawal, A., & Hockerts, K. (2019). Impact investing strategy: Managing conflicts between impact investor and investee social enterprise. *Sustainability*, 11(15), 4117.3. <https://doi.org/10.3390/su11154117>

Höchstädter, A. K., & Scheck, B. (2015). What's in a name: An analysis of impact investing understandings by academics and practitioners. *Journal of Business Ethics*, 132(2), 449–475.

Bell, B., & Haugh, H. (2015, April). Exploring institutional field emergence: Insights from social investment. In *SSFII Conference, Oxford* (pp. 23–25).

Bengo, I., & Calderini, M. (2016). New development: Are social impact bonds (SIBs) viable in Italy? A new roadmap. *Public Money & Management*, 36(4), 303–306.

Bhatt, P., & Ahmad, A. J. (2017). Financial social innovation to engage the economically marginalized: insights from an Indian case study. *Entrepreneurship and Regional Development*, 29(5-6), 391–413.

Bin Syed Azman, S. M. M., & Ali, E. R. A. E. (2016). Potential role of social impact bond and socially responsible investment sukuk as financial tools that can help address issues of poverty and socio-economic insecurity. *Intellectual Discourse*, 24.

Brandstetter, L., & Lehner, O. M. (2015). Opening the market for impact investments: The need for adapted portfolio tools. *Entrepreneurship Research Journal*, 5(2), 87–107.

Broccardo, E., Mazzuca, M., & Frigotto, M. L. (2020). Social impact bonds: The evolution of research and a review of the academic literature. *Corporate Social Responsibility and Environmental Management*, 27(3), 1316–1332.

Carè, R., & De Lisa, R. (2019). Social impact bonds for a sustainable welfare state: The role of enabling factors. *Sustainability*, 11(10), 2884. <https://doi.org/10.3390/su12187291>

Chen, S., & Harrison, R. (2020). Beyond profit vs. purpose: Transactional-relational practices in impact investing. *Journal of Business Venturing Insights*, 14, e00182.

Chiodo, V., Calderini, M., Arena, M., & Bengo, I. (2015). Social impact bond: New finance or new procurement? *ACRN Oxford Journal of Finance and Risk Perspectives*, 4(4), 168–189.

Clarkin, J. E., & Cangioni, C. L. (2016). Impact investing: A primer and review of the literature. *Entrepreneurship Research Journal*, 6(2), 135–173.

Cooper, C., Graham, C., & Himick, D. (2016). Social impact bonds: The securitization of the homeless. *Accounting, Organizations and Society*, 55, 63–82.

De Amicis, L., Binenti, S., Cardoso, F. M., Gracia-Lázaro, C., Sánchez, Á., & Moreno, Y. (2020). Understanding drivers when investing for impact: an experimental study. *Palgrave Communications*, 6(1), 1–9.

Espinosa, S. (2018). From philanthropy to impact investing: The case of Luxembourg. *Regions and Cohesion*, 8(1), 1–24.

FitzGerald, C., Carter, E., Dixon, R., & Airoidi, M. (2019). Walking the contractual tightrope: A transaction cost economics perspective on social impact bonds. *Public Money & Management*, 39(7), 458–467.

Fraser, A., Tan, S., Boaz, A., & Mays, N. (2020). Backing what works? Social impact bonds and evidence-informed policy and practice. *Public Money & Management*, 40(3), 195–204.

(continued)

- Glänzel, G., & Scheuerle, T. (2016). Social impact investing in Germany: Current impediments from investors' and social entrepreneurs' perspectives. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 27(4), 1638–1668.
- Hajer, J. (2020). The national governance and policy context of social impact bond emergence: A comparative analysis of leaders and skeptics. *Journal of Comparative Policy Analysis: Research and Practice*, 22(2), 116–133.
- Han, J., & Shah, S. (2019). The ecosystem of scaling social impact: A new theoretical framework and two case studies. *Journal of Social Entrepreneurship*, 11(2), 215–239.
- Hehenberger, L., Mair, J., & Metz, A. (2019). The assembly of a field ideology: An idea-centric perspective on systemic power in impact investing. *Academy of Management Journal*, 62(6), 1672–1704.
- Johnson, K. (2014). The charitable deduction games: Mimicking impact investing. *University of Pennsylvania Journal of Business Law*, 17(4), 1257. <https://scholarship.law.upenn.edu/jbl/vol17/iss4/5>
- Karpenko, L., & Voronzhak, P. (2018, May). Base alternatives and the paradigm of impact investing development in the coordinates of globalization changes and euro integration. In *Proceedings of the 4th International Conference on European Integration* (pp. 659–668).
- Kimbu, A. N., & Tichaawa, T. M. (2020). Determinants of impact investing for tourism development in emerging destinations of sub-Saharan Africa. *Development Southern Africa*, 37(5)1–16.
- Kish, Z., & Fairbairn, M. (2018). Investing for profit, investing for impact: Moral performances in agricultural investment projects. *Environment and Planning A: Economy and Space*, 50(3), 569–588.
- La Torre, M., Trotta, A., Chiappini, H., & Rizzello, A. (2019). Business models for sustainable finance: The case study of social impact bonds. *Sustainability*, 11(7), 1887. <https://doi.org/10.1016/j.jbvi.2019.e00141>
- Lehner, O. M., Harrer, T., & Quast, M. (2019). Building institutional legitimacy in impact investing. *Journal of Applied Accounting Research*, 20(3), 416–438.
- Lilley, S., Harvie, D., Lightfoot, G., & Weir, K. (2020). Using derivative logic to speculate on the future of the social investment market. *Journal of Urban Affairs*, 42(6), 920–936.
- Mayer, J., & Scheck, B. (2018). Social investing: What matters from the perspective of social enterprises? *Nonprofit and Voluntary Sector Quarterly*, 47(3), 493–513.
- McCallum, S., & Viviers, S. (2020). Private sector impact investment in water purification infrastructure in South Africa: A qualitative analysis of opportunities and barriers. *Water SA*, 46(1), 44–54.
- Michelucci, F. V. (2017). Social impact investments: does an alternative to the Anglo-Saxon paradigm exist? *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 28(6), 2683–2706.
- Muñoz, P., & Kimmitt, J. (2019). A diagnostic framework for social impact bonds in emerging economies. *Journal of Business Venturing Insights*, 12. <https://doi.org/10.1016/j.jbvi.2019.e00141>
- Ngoasong, M., Paton, R., & Korda, A. (2015). Impact investing and inclusive business development in Africa: A research agenda. The Open University, Milton Keynes.
- Ormiston, J., Charlton, K., Donald, M. S., & Seymour, R. G. (2015). Overcoming the challenges of impact investing: Insights from leading investors. *Journal of Social Entrepreneurship*, 6(3), 352–378.
- Ormiston, J., Moran, M., Castellás, E. I., & Tomkinson, E. (2020). Everybody wins? A discourse analysis of competing stakeholder expectations in social impact bonds. *Public Money & Management*, 40(3), 237–246.

(continued)

-
- Park, S. K. (2018). Social bonds for sustainable development: A human rights perspective on impact investing. *Business and Human Rights Journal*, 3(2), 233–255.
-
- Rania, F., Trotta, A., Carè, R., Migliazza, M. C., & Kabli, A. (2020). Social uncertainty evaluation of social impact bonds: A model and practical application. *Sustainability*, 12(9), 3854. <https://doi.org/10.3390/su12093854>
-
- Reisman, J., Olazabal, V., & Hoffman, S. (2018). Putting the “impact” in impact investing: The rising demand for data and evidence of social outcomes. *American Journal of Evaluation*, 39(3), 389–395.
-
- Roundy, P. T. (2019). Regional differences in impact investment: A theory of impact investing ecosystems. *Social Responsibility Journal*, 16(4).
-
- Scheck, B., Höchstädter, A. K., & Busch, T. (2016). Making money at the expense of the poor? An investigation of individuals’ preferences to impact invest versus donate. *Risk*, 5, 141–163.
-
- Schincus, C. (2017). Financial innovation as a potential force for a positive social change: The challenging future of social impact bonds. *Research in International Business and Finance*, 39, 727–736.
-
- Sinclair, S., McHugh, N., & Roy, M. J. (2019). Social innovation, financialisation and commodification: A critique of social impact bonds. *Journal of Economic Policy Reform*, 1–17.
-
- Trelstad, B. (2016). Impact investing: A brief history. *Capitalism & Society*, 11(2).
-
- Viviani, J. L., & Maurel, C. (2019). Performance of impact investing: A value creation approach. *Research in International Business and Finance*, 47, 31–39.
-
- Wilson, R., Fraser, A., Kimmitt, J., Tan, S., McHugh, N., Lowe, T., . . . & Carter, E. (2020). Theme: Futures in social investment? Learning from the emerging policy and practice of social impact bonds (SIBs) Editorial: Whither Social Impact Bonds (SIBs): The future of social investment? *Public Money and Management*, 40(3). <https://doi.org/10.1080/09540962.2020.1714287>
-
- Wong, M. C. S., & Yap, R. C. Y. (2019). Social impact investing for marginalized communities in Hong Kong: Cases and issues. *Sustainability*, 11(10), 2831. <https://doi.org/10.3390/su1102831>
-
- Reisman, J., Olazabal, V., & Hoffman, S. (2018). Putting the “impact” in impact investing: The rising demand for data and evidence of social outcomes. *American Journal of Evaluation*, 39(3), 389–395.
-
- Roundy, P. T. (2019). Regional differences in impact investment: A theory of impact investing ecosystems. *Social Responsibility Journal*, 16(4). <https://doi.org/10.1108/SRJ-11-2018-0302>.
-
- Mogapi, E. M., Sutherland, M. M., & Wilson-Prangley, A. (2019). Impact investing in South Africa: managing tensions between financial returns and social impact. *European Business Review*, 31(3)
-
- Karpenko, L., & Voronzhak, P. (2018, May). Base alternatives and the paradigm of impact investing development in the coordinates of globalization changes and euro integration. In *Proceedings of the 4th International Conference on European Integration* (pp. 659–668).
-
- Viviani, J. L., & Maurel, C. (2019). Performance of impact investing: A value creation approach. *Research in International Business and Finance*, 47, 31–39.
-
- Public Policy**
-
- Carter, E. (2019). More than marketised? Exploring the governance and accountability mechanisms at play in social impact bonds. *Journal of Economic Policy Reform*, 1–17.
-
- Choi, Y. J., Huber, E., Kim, W. S., Kwon, H. Y., & Shi, S. J. (2020). Social investment in the knowledge-based economy: New politics and policies. *Policy and Society*, 39(2), 147–170.
-
- Harvie, D., & Ogman, R. (2019). The broken promises of the social investment market. *Environment and Planning A: Economy and Space*, 51(4), 980–1004.
-
- Joy, M., & Shields, J. (2018). Austerity in the making: Reconfiguring social policy through social impact bonds. *Policy & Politics*, 46(4), 681–695.
-

(continued)

-
- Kim, J. (2018). Social finance funding model for animal shelter programs: Public-private partnerships using social impact bonds. *Society & Animals*, 26(3), 259–276.
- Lowe, T., Kimmitt, J., Wilson, R., Martin, M., & Gibbon, J. (2019). The institutional work of creating and implementing social impact bonds. *Policy & Politics*, 47(2), 353–370.
- Pandey, S., Cordes, J. J., Pandey, S. K., & Winfrey, W. F. (2018). Use of social impact bonds to address social problems: Understanding contractual risks and transaction costs. *Nonprofit Management and Leadership*, 28(4), 511–528.
- Phillips, S. D., & Johnson, B. (2019). Inching to impact: The demand side of social impact investing. *Journal of Business Ethics*, 1–15.
- Saltman, K. J. (2017). The promise and realities of pay for success/social impact bonds. *Education Policy Analysis Archives*, 25, 59.
- Smeets, D. J. A. (2017). Collaborative learning processes in social impact bonds: A case study from the Netherlands. *Journal of Social Entrepreneurship*, 8(1), 67–87.
- Tekula, R., & Andersen, K. (2019). The role of government, nonprofit, and private facilitation of the impact investing marketplace. *Public Performance & Management Review*, 42(1), 142–161.
- Temple, J. A., & Reynolds, A. J. (2015). Using benefit-cost analysis to scale up early childhood programs through pay-for-success financing. *Journal of Benefit-Cost Analysis*, 6(3), 628–653.
- Tse, A. E., & Warner, M. E. (2020). A policy outcomes comparison: Does SIB market discipline narrow social rights? *Journal of Comparative Policy Analysis: Research and Practice*, 22(2), 134–152.
- Warner, M. (2015). Profiting from public value? The case of social impact bonds. In J. M. Bryson, B. C. Crosby, & L. Bloomberg (Eds.), *Creating public value in practice: Advancing the common good in a multi-sector, shared-power, no-one-wholly-in-charge world* (pp. 143–160). Routledge.
- Social Policy**
-
- Berndt, C., & Wirth, M. (2018). Market, metrics, morals: The social impact bond as an emerging social policy instrument. *Geoforum*, 90, 27–35.
- Cefalo, R., & Kazepov, Y. (2018). Investing over the life course: The role of lifelong learning in a social investment strategy. *Studies in the Education of Adults*, 50(2), 203–218.
- Chiapello, E., & Knoll, L. (2020). The welfare conventions approach: A comparative perspective on social impact bonds. *Journal of Comparative Policy Analysis: Research and Practice*, 22(2), 100–115.
- Dowling, E. (2017). In the wake of austerity: Social impact bonds and the financialisation of the welfare state in Britain. *New Political Economy*, 22(3), 294–310.
- Edmiston, D., & Nicholls, A. (2018). Social impact bonds: The role of private capital in outcome-based commissioning. *Journal of Social Policy*, 47(1), 57–76.
- Findlay, S., & Moran, M. (2019). Purpose-washing of impact investing funds: motivations, occurrence and prevention. *Social Responsibility Journal*, 15 (7), 853–873.
- FitzGerald, C., Carter, E., Dixon, R., & Airoidi, M. (2019). Walking the contractual tightrope: A transaction cost economics perspective on social impact bonds. *Public Money & Management*, 39 (7), 458–467.
- FitzGerald, C., Fraser, A., & Kimmitt, J. (2020). Tackling the big questions in social impact bond research through comparative approaches. *Journal of Comparative Political Analysis*, 22 (2), 1–15.
- Fleckenstein, T., & Lee, S. C. (2019). Roads and barriers towards social investments: Comparing labour market and family policy reforms in Europe and East Asia. *Policy and Society*, 1–18.
- Fraser, A., Tan, S., Lagarde, M., & Mays, N. (2018). Narratives of promise, narratives of caution: A review of the literature on social impact bonds. *Social Policy & Administration*, 52(1), 4–28.
-

(continued)

-
- Hevenstone, D., & von Bergen, M. (2020). Public–private partnerships in social impact bonds: facilitating competition or hindering transparency? *Public Money & Management*, 40(3), 205–212.
-
- Kajiita, R. M., & Murote kang’ethe, S. (2020). Investing in people and communities for social change: Lessons from social enterprises in South Africa. *Journal of Progressive Human Services*, 1–18.
-
- Knoll, L. (2018). Social impact bonds and the welfare dispute. *Kolner zeitschrift fur soziologie und sozialpsychologie*, 70, 117–138.
-
- Maier, F., Barbetta, G. P., & Godina, F. (2018). Paradoxes of social impact bonds. *Social Policy & Administration*, 52(7), 1332–1353.
-
- Neyland, D. (2018). On the transformation of children at-risk into an investment proposition: A study of social impact bonds as an anti-market device. *The Sociological Review*, 66(3), 492–510.
-
- Ogman, R. (2019). “Ethical capitalism” in the city: Embedded economy or marketization? The case of social impact bonds. *Journal of Urban Affairs*, 1–23.
-
- Quinn, Q. C., & Munir, K. A. (2017). Hybrid categories as political devices: The case of impact investing in frontier markets. In R. Durand, N. Granqvist, A. Tyllström (Eds.) *From categories to categorization: Studies in sociology, organizations and strategy at the crossroads* (pp. 113–150). Emerald Publishing Limited.
-
- Rowe, R., & Stephenson, N. (2016). Speculating on health: Public health meets finance in ‘health impact bonds’. *Sociology of Health & Illness*, 38(8), 1203–1216.
-
- Riot, E. (2020). Betting on children at risk: Actors’ views on the introduction of social impact bonds in the French field of “childhood in danger”. *Journal of Urban Affairs*, 42(6), 856–891.
-
- Rizzello, A., & Kabli, A. (2020). Sustainable financial partnerships for the SDGs: The case of social impact bonds. *Sustainability*, 12(13), 5362.
-
- Roy, M. J., McHugh, N., & Sinclair, S. (2017). Social impact bonds—evidence-based policy or ideology? In B. Greve (Ed.), *Handbook of social policy evaluation* (pp. 263–275). Edward Elgar Publishing.
-
- Saltman, K. J. (2017). The promise and realities of pay for success/social impact bonds. *Education Policy Analysis Archives*, 25 (59). <https://doi.org/10.14507/epaa.25.2640> T
-
- Tse, A. E., & Warner, M. E. (2018). The razor’s edge: Social impact bonds and the financialization of early childhood services. *Journal of Urban Affairs* 42 (6), 816–832.
-
- Williams, J. W. (2018). Surveying the SIB economy: Social impact bonds, “local” challenges, and shifting markets in urban social problems. *Journal of Urban Affairs*, 1–13.
-

Health

- Abdullah, F., Naledi, T., Nettleship, E., Davids, E. L., Van Leeuw, L., Shangase, S., . . . & Mathews, C. (2019). First social impact bond for the SAMRC: A novel financing strategy to address the health and social challenges facing adolescent girls and young women in South Africa. *South African Medical Journal*, 109(11b), 57–62.
-
- Jones, C., Hartfield, N., Brocklehurst, P., Lynch, M., & Edwards, R. T. (2020). Social return on investment analysis of the health precinct community hub for chronic conditions. *International Journal of Environmental Research and Public Health*, 17(14), 5249.
-
- Torres, E. M., Fernández, Ó. F., Rueda, P. C., Ruiz-Beato, E., Pérez, E. E., Estrada, R. M., . . . & Merino, M. (2020). Social value of a set of proposals for the ideal approach of multiple sclerosis within the Spanish National Health System: A social return on investment study. *BMC Health Services Research*, 20(1), 84.
-

Political Economy

- Andreu, M. (2018). A responsibility to profit? Social impact bonds as a form of “humanitarian finance”. *New Political Science*, 40(4), 708–726.
-

(continued)

Ryan, S., & Young, M. (2018). Social impact bonds: The next horizon of privatization. *Studies in Political Economy*, 99(1), 42–58.

Geography

Cohen, D., & Rosenman, E. (2020). From the school yard to the conservation area: Impact investment across the nature/social divide. *Antipode*, 52(5), 1259–1285.

References

- Adelman, I. 2000. The role of government in economic development. In *Foreign aid and development: Lessons learnt and directions for the future*, ed. F. Tarp, 53–75. London: Routledge.
- Broccardo, E., and M. Mazzuca. 2019. The missing link? Finance, public services, and co-production: The case of social impact bonds (SIBs). *Public Money & Management* 39 (4): 262–270.
- Bar-Mor, H. 1999. *The third sector: Legal aspects*. Israeli Center for Third Sector Research, Ben-Gurion University of the Negev, Beersheba. (in Hebrew)
- Calipha, R., and B. Gidron. 2020. The evolution of the Israeli third sector: A conceptual and empirical analysis. *Voluntaristics Review* 5 (4).
- Clarkin, J.E., and C.L. Cangioni. 2016. Impact investing: A primer and review of the literature. *Entrepreneurship Research Journal* 6 (2): 135–173.
- Daggers, J., and A. Nicholls. 2016. *The landscape of social impact investment research: Trends and opportunities*. Said Business School: Oxford University.
- Eran, Y. 1992. *Voluntary organizations in the welfare system*. Tel Aviv: Ramot.
- EVPA. 2017. Investing for impact, September 15 2017, GSG Impact Summit 2017, in <https://evpa.eu.com/blog/gsg-impact-summit-2017>
- Florida, R., G. Gates, B. Knudsen, and K. Stolarick. 2006. *The university and the creative economy*. <http://creativeclass.com/rfcgdb/articles/University%20For%20City%20and%20Community%204.pdf>
- Friedrich-Silber, I., and Z. Rosenhek. 2000. *The historical development of the Israeli third sector*. Beer-Sheva: Ben-Gurion University of the Negev.
- Galnoor, I. 2003. The judicialization of the public sphere in Israel. *Israel Law Review* 37: 500.
- Geiger, R.L. 1986. *Private sectors in higher education: Structure, function, and change in eight countries*. Ann Arbor: University of Michigan Press.
- Geiger, R. 1993. *Research and relevant knowledge*. New York: Oxford University Press.
- Gidron, B. 1984. Predictors of perseverance and dropout at work among community center volunteers. *Magamot* 29 (2): 180–198. (in Hebrew).
- . 1997. The evolution of Israel's third sector: The role of predominant ideology. *Voluntas: International Journal of Voluntary and Nonprofit Organizations* 8 (1): 11–38.
- . 1999. *Database of the third sector in Israel: Preliminary findings: Registry of associations 1981–1998, government support for third sector organizations 1991–1997*. Israeli Center for Third Sector Research, Ben-Gurion University of the Negev, Beersheva. (in Hebrew)
- Gidron, B., and D. Bargal. 1986. Self-help awareness in Israel: An expression of structural changes and expanding citizen participation. *Journal of Voluntary Action Research* 15 (2): 47–56.
- Gidron, B., and M. Chesler. 1994. Universal and particular attributes of self-help: A framework for international and intranational analysis. In *Self-help and mutual aid groups*, ed. F. Lavoie, T. Borkman, and B. Gidron, 1–44. New York: The Haworth Press.
- Gidron, B., and H. Katz. 1998. *The third sector in Israel. First analysis: boundaries, characters and processes*. Israeli Center for Third Sector Research, Ben-Gurion University of the Negev, Beersheva. (in Hebrew)

- Gidron, B., M. Bar, and H. Katz. 2003. The Israeli third sector: Between welfare state and civil society. Hakibbutz Hameuchad, Israel. (in Hebrew)
- . 2004. *The Israeli third sector*. Springer US, Kluwer Academic, New York
- GIIN. 2016. *2016 Annual impact investor survey*. <https://thegiin.org/research/publication/annualsurvey2016>
- . 2020. *2020 Annual impact investor survey*. <https://www.thegiin.org/research/publication/impinv-survey-2020>
- GSG. 2019. *Transition of impact economies: A global overview*. <https://www.gsgii.org/reports/transition-of-impact-economies-a-global-overview/>
- GSG. 2021. G7 Impact Taskforce announces membership, in <https://gsgii.org/2021/08/g7-impact-taskforce-announces-membership/>
- Höchstädter, A. K., & Scheck, B. 2015. What's in a name: An analysis of impact investing understandings by academics and practitioners. *Journal of Business Ethics* 132(2): 449–475.
- Jaffe, E.D. 1979. Non-conventional philanthropy. *Moment Magazine* 5 (5): 63–64.
- . 1982. *Giving wisely: The Israel guide to volunteer and nonprofit social services*. Jerusalem: Koren Publishers.
- . 1983. *Claim and protest: Future prediction of voluntary organizations in Israel*. Jerusalem: American Jewish Committee. (in Hebrew).
- . 1985. *Givers and spenders: The politics of charity in Israel*. Jerusalem: Ariel Publishers.
- . 1987. The crisis in Jewish philanthropy. *Tikkun Magazine* 2 (4): 27–31.
- Jamieson, D., R. Wilson, M. Martin, T. Lowe, J. Kimmitt, J. Gibbon, and M. French. 2020. Data for outcome payments or information for care? A sociotechnical analysis of the management information system in the implementation of a social impact bond. *Public Money & Management* 40 (3): 213–224.
- Karpenko, L., & Voronzhak, P. 2018. Base alternatives and the paradigm of impact investing development in the coordinates of globalization changes and euro integration. In Proceedings of the 4th International Conference on European Integration (pp. 659–668).
- Katan, Y. 1988. Voluntary organizations: Substitute or partner of the state in a welfare field. *Social Security* 32: 57–73. (in Hebrew).
- Kramer, R.M. 1970. *Community development in Israel and the Netherlands: A comparative analysis*. Berkeley: Institute of International Studies, University of California.
- . 1976. *The voluntary service agency in Israel*. Berkeley: Institute of International Studies, University of California.
- . 1981. *Voluntary agencies in welfare state*. Berkeley: University of California Press.
- . 1984. Voluntary agencies and social change in Israel, 1972–1982. *Israel Social Science Research* 2 (2): 55–72.
- Kubátová, J., and O. Kročil. 2020. The potential of impact and integral investing for sustainable social development and the role of academia in their dissemination. *Sustainability* 12 (17): 6939.
- Leslie L, and J.S. Caldwell. 1990. *Qualitative reading inventory*. New York: Harper Collins.
- Leslie, S.W. 1993. *The Cold War and American science: The military-industrial-academic complex at MIT and Stanford*. New York: Columbia University Press.
- Lyons, T.S., and J.R. Kickul. 2013. The social enterprise financing landscape: The lay of the land and new research on the horizon. *Entrepreneurship Research Journal* 3 (2): 147–159.
- Mersland, R., Nyarko, S. A., & Sirisena, A. B. 2020. A hybrid approach to international market selection: The case of impact investing organizations. *International Business Review* 29 (1):101624. <https://doi.org/10.1016/j.ibusrev.2019.101624>
- Nicholls, A., and C. Pharoah. 2008. *The landscape of social investment: A holistic topology of opportunities and challenges*. https://www.sbs.ox.ac.uk/sites/default/files/2019-10/Landscape_of_Social_Investment.pdf
- OECD. 2019. *Social impact investment 2019: The impact imperative for sustainable development*. Author. <https://www.oecd.org/development/social-impact-investment-2019-9789264311299-en.htm>

- Reisman, J., Olazabal, V., & Hoffman, S. 2018. Putting the “impact” in impact investing: The rising demand for data and evidence of social outcomes. *American Journal of Evaluation* 39(3): 389–395.
- Rizzello, A., Migliazza, M., Care, R., & Trotta, A. 2016. A model and research agenda. Routledge handbook of social and sustainable finance 102.
- Salamon, L.M., H.K. Anheier, R. List, S. Toepler, S.W. Sokolowsky, and Associates. 1999. *Global civil society: Dimensions of the nonprofit sector*. Baltimore: The Johns Hopkins University Institute for Policy Studies.
- Shay, S., A. Lazar, R. Duchin, and B. Gidron. 1999. *Philanthropy in Israel: Patterns of public donation and volunteering – research report*. Israeli Center for Third Sector Research, Ben-Gurion University of the Negev, Beersheva. (in Hebrew).
- Spiess-Knafl, W., and J. Aschari-Lincoln. 2015. Understanding mechanisms in the social investment market: What are venture philanthropy funds financing and how? *Journal of Sustainable Finance & Investment* 5 (3): 103–120.
- Statista. 2020. <https://www.statista.com/statistics/183995/us-college-enrollment-and-projections-in-public-and-private-institutions/>
- Stiglitz, J.E. 1997. The role of government in economic development. In *Annual World Bank conference on development economics*, vol. 1996, 11–23. Washington, DC: World Bank.
- Telias, M., J. Katan, and B. Gidron. 2000. *Government and local authority policy towards the third sector in Israel*. Israeli Center for Third Sector Research, Ben-Gurion University of the Negev, Beersheva. (in Hebrew)
- UNCTAD. 2020. World investment report 2020: International production beyond the pandemic. United Nations New York <https://unctad.org/webflyer/world-investment-report-2020>.
- Yishai, Y. 1987. *Land or peace: Whither Israel?* Vol. 352. Stanford: Hoover Institution Press.



Financial Crisis, Capital Requirement, and Stress Tests: Evidence from the Extreme Value and Stable Paretian Estimates

85

Tony Sio-Chong U and Jacky Yuk-Chow So

Abstract

Stress tests are exercises to determine the losses that might occur under “unlikely but plausible” circumstances. The objective of this chapter is to demonstrate that the “fat tail” of the stable Paretian distribution can capture the extraordinary risk during financial crisis, as well as normal market conditions. Rather than assuming a normal distribution and then add the “stress tests,” a stable Paretian VaR will capture both and is more parsimonious statistically. The Central Limit Theorem and probability theory are applicable to the “stress tests.” (Diebold et al., *J Risk Finance* 1(2):30–35, 2000) believe that extreme value statistics are preferable. This argument is valid if regulators and corporations are interested in managing “extreme risk” only. The soundness and the safety of the financial markets and the banking system and the “externality” issues, however, reveal that good risk management must consider “normal market” conditions as well as “crisis” that will be captured by the density of the whole distribution, not just the tails. In this chapter, a relatively new technique will be used to choose the appropriate number of “extremes.” The difficulty in estimating the density of the stable Paretian distribution will be overcome by using a simple and reliable Fourier series method.

Keywords

Financial crisis · Adequacy capital ratio · Tail risk · VaR · Fourier-series

JEL Classification

C46 · C63 · G01 · G28 · G33

T. S.-C. U · J. Y.-C. So (✉)

Department of Accounting and Finance, Macau University of Science and Technology, Taipa, Macau

e-mail: scu@must.edu.mo; jycso@must.edu.mo

85.1 Introduction

From 2008 to 2012, the financial crisis in the USA and Europe highlighted the limitation of the Value-at-Risk (VaR) method as an effective risk management tool for financial managers and regulators. Because VaR uses a backward-looking approach and assumes a normal market and a normal distribution, “extra-ordinary” risk created by financial crises and “non-normal” fat-tailed distribution cannot be captured accurately. In other words, portfolio risk, and thus capital for banks and firms, will be insufficient to cover the extremely large losses that are underestimated by a normal distribution. Akhigbe et al. (2012) and Garel and Petit-Romec (2017) showed that bank capital is necessary but may not be sufficient to prevent crisis. In addition, the FDIC insured the deposits as the failure of their insured banks.¹ The regulation seems safe and sound for both banks and households. The reality is a billion dollars are injected into the banking system. Most financial institutions request government’s bailout. Buser et al. (1981) have showed that the FDIC might have intentionally underpriced the insurance premium to attract more banks to be FDIC members. Besides, the banks have underestimated the risk; the fragility of financial system is completely exposed.

Stress tests have been recommended by the regulators, such as IMF, Bank of International Settlements (BIS), US Federal Reserve Banks, and the European Banking Authority (EBA), etc., to “complement” the results from the VaR analyzes since the Asian financial crisis and the bankruptcy of the Long Term Capital Management (LTCM) Ltd. The usefulness and the effectiveness of stress tests, however, are not subject to serious testing until the recent global financial crisis triggered by the subprime mortgage loan crisis in the USA.

The principle of stress tests is to determine the losses that might occur due to the bad state of economy. After conducting the stress tests, US bank regulators announced that ten big US banks must raise \$75 billion. Money center banks such as the Bank of America, Citigroup, Wells Fargo, etc., were required to increase their Supervisory Capital Assessment Program (SCAP) buffer by \$33.9 billion, \$5.5 billion, and \$13.7 billion, respectively. Financial company, such as GMAC, was also required to increase capital by \$11.5 billion.

In 2011, the European Banking Authority conducted similar stress test for 90 banks in 21 countries and found that eight European banks are not strong enough to withstand a prolonged recession and need to raise €2.5 billion in capital. German bank Helaba failed the test but withdrew from the test, contesting the methodology used and releasing its own results on its website (see BBC News, Business, July 15, 2011). Nordbank and Norddeutsche Landesbank both challenged the stress test results saying they didn’t reflect how strong they were, with Nordbank saying that “this result in no way reflects the economic reality of our banks” (“8 Banks Flunks European Stress Test; 16 More Barely Pass,” USA Today, July 15, 2011). The Greek

¹The standard insurance amount is \$250,000 per depositor, per insured bank, for each account ownership category.

finance minister, Evangelos Venizelos, said that “the criteria were particularly harsh, they were extreme.”

Obviously, regulators believe that stress tests are important for the soundness of financial markets and institutions. However, some corporate executives have reservation about the stress test results. The reservation may be due to the “externality” created by the stress test results and the statistical deficiency of the stress tests.

1. Externality and Signal: Mr. William Dudley, President and CEO of the Federal Reserve Bank of New York believes that banks do not want to proactively raise capital during the financial crisis because capital raises might signal that they were in fact weak and stock price may be depressed.² Bank’s decision not to raise capital increases the likelihood of the “bad state of the world” – an important externality. Additional capital doesn’t just make the bank stronger, it also makes the entire banking system more stable. This is the collective action problem that the SCAP was designed to remedy.
2. Deficiency of stress tests: many researchers have pointed out that the stress tests are ad hoc and not probabilistic. Without an idea of the related probability estimates, the top management does not know whether to treat the results seriously or not. Raising a large amount of stress capital will increase the cost of capital significantly.

The objective of this study is to conduct empirical tests to demonstrate that, similar to the VaR plus stress test, the “fat tail” of the stable Paretian distribution can capture extraordinary risk during financial crisis. Since the normal distribution is a special case of the stable Paretian distribution, “normal market” condition is implicitly captured. In addition, because the stable distribution belongs to the stable family, central limit theorem (CLT) holds and probability can be assigned to “stress tests.” With the more accurate estimates for the worst-case scenario derived from the stable family, risk-based capital can be estimated for different banks or companies that have different risk. This risk-based capital system will reduce the externality problem.

Because the density of the stable Paretian, based on characteristic function, is quite challenging, this chapter proposes a Fourier series-based method to calculate Value-at-Risk (VaR). This method can be used to evaluate the VaR of portfolio with different characteristic exponents (α) of assets; the key idea is to expand the distribution with Fourier series and the coefficients of the expansion are close relatively to the characteristic functions of returns of different assets. Note that Samuelson (1967) and Merton (1969) have studied portfolio risk for symmetric stable Paretian distribution. In order to capture financial crisis, we will use the “rare-event” equilibrium model recommended by Liu et al. (2005).

Our preliminary empirical results show that the VaR plus stress test can be more accurately estimated by using the stable Paretian VaR. For normal distribution, the risk of assets is determined by one factor, volatility (σ) and the upside and downside

²www.newyorkfed.org/newsevents/speeches/2011/dudl

movements of returns are symmetric. Therefore, it cannot predict the catastrophic market risk precisely. For stable Paretian distribution, the risk of assets is determined by three factors; characteristic exponent (α), skewness (β), and scale (c). The catastrophic market risk or extraordinary risk is captured by characteristic exponent (α), skewness (β) values the asymmetric movements of returns, and the ordinary risk is predicted by scale (c). Moreover, normal distribution is the special case of stable Paretian distribution with $\alpha = 2$ and $\beta = 0$. The tail index (characteristic exponent) is critical; DuMouchel (1983) let the tails tell the story also assumes that the appropriate number of extremes can be determined. McCulloch (1997) has pointed out the importance of large sample size for testing the “stability” of the stable family. The comparison between normal distribution, stable Paretian distribution with Coronel-Brizio and Hernandez-Montoya (2005)’s extreme value method shows that the VaR estimations based on different α of assets yielded more accurate than other models, the detail is given in So and U (2012).

This chapter also makes two more contributions to the literature. First, we document that some companies met the required capital by increasing the weight of risk-free assets instead of raising capital. This asset allocation strategy is consistent with the fact that most banks prefer not to increase capital during the recession but using a low risk-low return investment strategy. This safety-first rule may be appropriate during financial crisis and is not because of the concern of weakness signaled nor its negative possible impact on stock prices. It will not create externality either because if the risk of the asset portfolio is low enough, no new capital will be necessary.³ Our results show that Bank of America, Wells Fargo, and Citigroup had inadequate capital in 2008 that is consistent with the announcement of US bank regulators. After that, we found that most banks increased their weight of risk-free assets to meet the required capital ratio instead of raising capital recommended by US bank regulators.

85.2 Literature Review

After the subprime mortgage crisis in 2008, many literatures have found that liquidity and credit risks are very critical for maintaining an orderly market as well as the confidence of investors. (see Castiglionesi et al. (2014), Cornett et al. (2011), Freixas et al. (2011), Ivashina and Scharfstein (2010), and Keister and McAndrews (2009) for more details on liquidity risk. See Correa et al. (2014), Galai and Wiener (2012), and Yang and Zhou (2013) for more details on credit risk.) These studies found that market liquidity substantially decreased in 2008. Banks then reacted by holding excess reserves, which resulted in higher borrowing costs and many corporations and households were not able to obtain financing from financial institutions. Consequently, the subprime mortgage crisis spread from the financial sector to the real economy. The

³Some European banks in fact held lot of government bonds after 2008 because they are classified as the safest assets and required no risk-based capital. That laid down the European Debt crises when the sovereign risks of some countries were very high.

latter was a key factor that escalated the propagation of the subprime mortgage crisis. The credit default swap spreads increased to 1500 basis points (Ivashina and Scharfstein 2010), and the credit ratings of most banks in the United States were downgraded. Their stock prices plunged more than 30%. These significant losses exacerbated the liquidity freeze in the real economy. Allen and Gale (2001) developed a model to explain how bubbles were created by risky assets and concluded that the economic crisis was the consequence of a bubble burst. However, the classic definition of the bubble should rely on the fundamental value of risky assets, it cannot be easily verified because the pricing methods of different assets (e.g., real estates, stocks, and bonds) are quite different from each other. In Allen and Gale (2001), the fundamental value of risky assets is the discounted value of the net returns.

Yet how could banking systems be so vulnerable to failure even though they are carefully monitored by regulators? Flannery (2014) estimates the probability of default for 25 of the largest banks in USA based on their market equity. The results show that the probability of their default far exceeded the 0.1% standardized risk weight stipulated by the Basel standard. He concludes that the supervisory metrics from book capital measures are seriously flawed. In other words, the tail risk measured by market value is much higher than that in the book value measures. Consequently, capital requirements facilitated by using an accounting-based ratio do not actually reveal the real risks of financial institutions. Hovakimian and Kane (2000) point out that bank accountants are prone to overstating the amount of capital and understating the risks, even so regulators seem reluctant to enforce market value-based measures instead of book value-based measures.

On the other hand, defining and measuring crises are challenging tasks. Allen and Gale (2001) document different crises over the past three decades. However, they find that it is difficult to define the length and depth of market downturns before they could be viewed as a “crisis.” Acharya et al. (2012) and Brownlees and Engle (2011) indicate that a proxy for financial crises would be a decline of over 40% over the next 6 months in the aggregate stock market. Therefore, a market value-based measure to evaluate the capital losses during the crisis is crucial and necessary. Here, a parametric method is recommended because it is more efficient. Instead of using simulations, the Fourier-cosine expansion is employed to produce more accurate estimates, especially for the stable Paretian distribution. VaR with fat-tailed distribution under an efficient numerical approach, thereby can provide a comprehensive risk measure for the practical use of banks internally and supervisory purposes.

85.3 Methodology

In this chapter, the integrated VaR and stress test method will be proposed. VaR is a financial measure that evaluates the return with a specific probability. The stress test is a risk management tool for measuring the impact on firm of specific events and/or movements in a set of financial variables. Accordingly, the stress test is viewed as a complement to the VaR model. In our model, we simply consider the values of total

assets⁴ (TA_{t+1}) are greater than or equal to the debts ($D_{t,t+1}$) at time t and the maturity at $t+1$ with the interest rate of debt (r_d), that is,

$$\begin{aligned} TA_{t+1} &\geq D_{t,t+1}(1+r_d) \\ TA_t(1+r_{t+1}) &\geq D_{t,t+1}(1+r_d) \\ \frac{1+r_{t+1}}{1+r_d} &\geq \frac{D_{t,t+1}}{TA_t}, \end{aligned} \quad (1)$$

where r_{t+1} are the returns of total assets at time $t+1$. Denote the debt per assets ratio (d) as $d = \frac{D_{t,t+1}}{TA_t}$, and capital ratio (cr),⁵ $cr = 1 - d$, then Eq. (1) becomes,

$$cr \geq \frac{r_d - r_{t+1}}{1 + r_d} \quad (2)$$

Our estimations of r_{t+1} are continuous compound returns,⁶ Eq. (2) expressed as

$$cr \geq 1 - \exp(r_{t+1} - r_d). \quad (3)$$

The estimations of r_{t+1} in recession are the critical part of our model.⁷

The total assets of firms can be viewed as a portfolio. The $w_{0t}, w_{1t}, \dots, w_{nt}$ are the weights of assets at time t , $r_{0t+1}, r_{1t+1}, \dots, r_{nt+1}$ are the continuous compound returns of each asset at time $t+1$. We specified r_{0t+1} be the risk-free rate r_f . The returns of the total assets or the portfolio (r_{pt+1}) at time $t+1$ are given by,

$$r_{pt+1} = w_{0t}r_f + w_{1t}r_{1t+1} + \dots + w_{nt}r_{nt+1}, \quad (4)$$

thus, the returns with a given probability θ can be determined by the VaR model. Therefore, Eq. (3) becomes

$$cr \geq 1 - \exp(VaR(\theta, w_{0t}, \dots, w_{nt}) - r_d), \quad (5)$$

where $VaR(\theta, w_{0t}, \dots, w_{nt})$ is given as follows,

$$\Pr(w_{0t}r_f + \dots + w_{nt}r_{nt+1} \leq VaR(\theta, w_{0t}, \dots, w_{nt})) = \theta. \quad (6)$$

We denoted the required capital ratios (rcr) as follows

⁴It contains all risky assets and a risk-free asset.

⁵In this study, we simply assume the debt is equal to the liability. Therefore, we have $TA = L + C = D + C$ and $cr + d = 1$. If the total liability included long term liability (LL), then we may add the long-term liability ratio ($l = LL/TA$) to the left-hand side of Eq. (2).

⁶ $r_{it+1} = \ln\left(\frac{p_{it+1}}{p_{it}}\right)$ and p_{it} are the price levels at time t .

⁷Regime shift model can generate mixed distribution that has fatter tail than normal. However, the CLT may not be directly applied to obtain the distribution of aggregate returns.

$$rcr = 1 - \exp(VaR(\theta, w_{0t}, \dots, w_{nt}) - r_d). \quad (7)$$

The right-hand side of Eq. (5) is called the required capital ratio.⁸ If the capital ratio is high enough, we could invest much more in the risky assets. However, companies invest too much in risky assets without adequate capital, such that Eq. (5) is invalid, the company will have a probability of θ or higher to go bankrupt, we denote this as the probability of bankruptcy or bankruptcy risk. In this case, the companies had two ways to avoid bankruptcy. One way is to increase the capital ratio, the other way is to increase the weight of risk-free assets, thus the VaR decreased and Eq. (5) can be fulfilled.⁹

In practice, companies needed to maximize the returns and also reach the required capital ratio. Therefore, the following equation should be satisfied,

$$\begin{aligned} & \max_{w_{0t}, \dots, w_{nt}} E[w_{0t}r_f + \dots + w_{nt}r_{nt+1}], \\ & \text{subject to } cr \geq 1 - \exp(VaR(\theta, w_{0t}, \dots, w_{nt}) - r_d). \end{aligned} \quad (8)$$

Using Lagrange multiplier method, we can solve (8).

85.4 VaR Models

In the previous section, the critical part is to determine the loss of the total asset with a specific probability θ . In this section, VaR models will be proposed, and $VaR(-\theta, w_{0t}, \dots, w_{nt})$ are evaluated by a Fourier series-based method. We considered the one factor model, let the returns of different assets at $t + 1$ (r_{it+1}) be

$$r_{it+1} = a_i + b_i X_{t+1} + \varepsilon_{it+1}, \quad (9)$$

where $X_{t+1} \sim S(\alpha_X, \beta_X, c_X, \delta_X)$ is the α_X -stable Paretian distribution,¹⁰ it is independent of the residuals ε_{it+1} and $\varepsilon_{it+1} \sim S(\alpha_i, \beta_i, c_i, \delta_i)$. Moreover, ε_{it+1} and ε_{jt+1} are statistically independent for any $i \neq j$. ε_{it_1} and ε_{it_2} are also statistically independent for all $t_1 \neq t_2$. Nolan and Ojeda-Revah (2013)'s algorithm can be used to estimate the

⁸Buser et al. (1981) showed optimal capital structure of banks with and without FDIC insurance and with "underpriced FDIC insurance premium." In contrast with their optimal capital structure, our required capital ratio focuses on the optimal capital structure under a very low probability of default rather than the break even point of tax shield and the cost of bankruptcy in Buser et al. (1981). Our optimal structure may be suboptimal in Buser et al. (1981) but more reliable in the whole economy.

⁹Acharya et al. (2014) pointed out that the bank bailouts may transfer the financial sector credit risk to sovereign credit risk. Consequently, the sovereign bonds became more risky. If risk free assets are not really risk free, such as the European sovereign bonds, the risk could be very high: two way causality of the European debt crisis in 2002. In this situation, we may allow the r_f follows another stable distribution.

¹⁰We consider X_{t+1} as the benchmark return, it is similar to the market return of Capital Assets Pricing Model.

parameters when the residuals are stable distribution. We can extend this model to Fama-French model (Fama and French 1993) or ARMA model. Therefore, the returns of total assets at time $t + 1$ can be expressed by,

$$r_{pt+1} = w_{0t}r_f + \sum_{i=1}^n w_{it}r_{it+1} = A_t + B_tX_{t+1} + \phi_{t+1}, \quad (10)$$

where $A_t = w_{0t}r_f + \sum_{i=1}^n w_{it}a_i$, $B_t = \sum_{i=1}^n w_{it}b_i$ and $\phi_t = \sum_{i=1}^n w_{it}\varepsilon_{it+1}$. If ε_{it+1} are independent stable Paretian distributed with same characteristic exponent, α_ε , then ϕ_{t+1} is a stable Paretian distribution. The proof is given by property 1 in McCulloch and Lee (2008),

$$\phi_{t+1} = \sum_{i=1}^n w_{it}\varepsilon_{it+1}, \quad (11)$$

we have

$$\phi_{t+1} \sim S(\alpha_\varepsilon, \beta_{\varepsilon t}, c_{\varepsilon t}, \delta_{\varepsilon t}),$$

where

$$\begin{aligned} \beta_{\varepsilon t} &= \frac{\sum_{i=1}^n |w_{it}|^\alpha c_i^\alpha \text{sign}(w_{it}) \beta_i}{c_{\varepsilon t}^\alpha}, \\ c_{\varepsilon t} &= \left(\sum_{i=1}^n |w_{it}|^\alpha c_i^\alpha \right)^{\frac{1}{\alpha}}, \quad \text{and} \\ \delta_{\varepsilon t} &= \sum_{i=1}^n w_{it} \delta_i. \end{aligned} \quad (12)$$

We call this model as Stable (SA). Moreover, the distribution of r_{pt+1} can be estimated by Theorem 1 with $m = 2$. If ε_{it+1} are independent stable Paretian distributions with different characteristic exponents, say α_i , then we need to use Theorem 1 with $m = n + 1$ to estimate the distribution of r_{pt+1} , and we call this model as Stable (DA).

Theorem 1 If Y_j are independent random variables, with $Y_j \sim S(\alpha_j, \beta_j, c_j, \delta_j)$ for all $j \in \mathbb{N}$, and the cumulative distribution functions of each Y_j can be approximated in a closed interval $[u, v]$, then the cumulative distribution function of $Z = \sum_{j=1}^m \omega_j Y_j$ can be approximated by following,

$$\Pr(Z < s) \approx \frac{s - u}{v - u} + \sum_{k=1}^N \frac{(b - a)\tilde{A}_k}{k\pi} \sin\left(k\pi \frac{s - u}{v - u}\right),$$

where

$$\tilde{A}_k = \frac{2}{b - a} \operatorname{Re} \left\{ \exp\left(-i \frac{k\pi}{b - a}\right) \prod_{j=1}^m \varphi_j\left(\frac{k\omega_j\pi}{b - a}\right) \right\},$$

$i = \sqrt{-1}$ and φ_j is the characteristic function of Y_j . The proof of this Theorem is given in U et al. (2016).

According to Theorem 1, we can estimate $\Pr(r_{pt+1} < y)$ in Stable (SA) with $Y_1 = X_{t+1}$ and $\omega_1 = B_t$, $Y_2 = \phi_{t+1}$ and $\omega_2 = 1$ and $s = y - A_t$. For the Stable (DA), we can set $Y_1 = X_{t+1}$ and $\omega_1 = B_t$, $Y_2 = \varepsilon_{j-1t+1}$ and $\omega_j = w_{j-1t}$ for $j = 2, \dots, n+1$ and $s = y - A_t$.

85.5 Empirical Results

85.5.1 Investment Companies

In this section, we considered an investment company that invested in two stocks, American Express (AXP) & Cisco System (CSCO), and a risk-free asset (r_f). The returns of Dow Jones Index are used as benchmark index (X_{t+1}) in Eq. (10). We chose weekly data from 2/1992 to 12/2010, we first used weekly data¹¹ from 2/1992 to 8/2001 and Nolan and Ojeda-Revah (2013)'s algorithm¹² to estimate a_i and b_i in Eq. (9). The Nolan (2001) maximum likelihood technique is used to estimate parameters in different models:

- The residuals in (9) are stable, Paretian distribution with different α , we say stable (DA).
- The residuals (ϕ_{t+1}) in Eq. (10) are stable Paretian distribution that is given by Eq. (11), we say stable (SA).
- The traditional one-factor model with normal uncertainty disturbance, we say normal.

And parameters of above models are given in Table 1. We then determine the VaR in Eq. (6) from 9/2001 to 12/2010 weekly data¹³ with Theorem 1.

In Tables 2 and 3, we assumed the interest rate of debt (r_d) in Eq. (5) was equal to zero. Table 2 shows that the required capital ratio with the probability of bankruptcy is less than 0.1% under different weights of risk-free assets. We use different models to compare the impact of the changes in allocation of assets on the required capital ratio. (a) Stable (DA) model: the required capital ratio was reduced rapidly if the company invested less in CSCO, but the required capital ratio was reduced slowly if we only decrease the weight of AXP. (b) Stable (SA) model: we take average of α of two assets; therefore, the extraordinary risk of these two assets is assumed to be the same. The result shows that difference in required capital ratios among different methods of reducing the weight of two risky assets is not obvious. The largest

¹¹ We called these as in the sample data.

¹² This algorithm allows the asymmetric stable disturbance in the regression.

¹³ We call these as out of the sample data.

Table 1 The description of parameters of residuals. In the stable Paretian distribution, Nolan (2001)’s Maximum Likelihood Estimation algorithm is used to estimate the parameters, the regression results in the stable Paretian distribution came from Nolan and Ojeda-Revah (2013). The stable distribution on the left-hand side and the normal distribution ($\alpha = 2$ and $\beta = 0$) on the right-hand side

Stable	X_{t+1}	AXP	CSCO	Normal	X_{t+1}	AXP	CSCO
a	–	0.002	0.0143	a	–	0.0005	0.0042
b	–	1.3589	1.2149	b	–	1.3619	1.2568
α	1.8021	1.9217	1.6676	σ	0.0204	0.0323	0.0595
β	0.0018	–0.5043	–0.5246	μ	0.0023	–	–
c	0.0128	0.0307	0.0445				
δ	0.0206	–	–				

Table 2 The capital ratio at different weights of risk-free assets with (1) reducing the same weights of two risky assets (equal), (2) reducing the weight of AXP and keeping CSCO as constant (rAXP), and (3) reducing the weight of CSCO and keeping AXP as constant (rCSCO), the starting point is 10% of risk-free assets, and each risky asset is 45%. The probability of bankruptcy θ is 0.1%. The unit of each cell is in percentage

	Capital Ratio (cr)								
	Stable (DA)			Stable (SA)			Normal		
w_0	Equal	rAPX	rCSCO	Equal	rAPX	rCSCO	Equal	rAPX	rCSCO
10	55.44	55.44	55.44	43.86	43.86	43.86	11.49	11.49	11.49
20	49.28	54.1	44.67	38.99	40.41	37.94	10.21	10.67	9.85
30	43.12	52.96	34.26	34.11	37.4	32.54	8.94	9.94	8.33
40	36.96	52.03	24.61	29.24	34.96	27.98	7.66	9.32	7.01
50	30.8	51.35	16.96	24.36	33.27	24.82	6.38	8.85	6.04

difference in capital ratio between rAXP and rCSCO is around 10%. It is much smaller than the largest difference in capital ratio (35%) in Stable (DA). In a normal model, the required capital ratio is very low, and its changes are not sensitive to the changes of the weight of risk-free assets. Figure 1 shows the graphic results of Table 2. If a company has inadequate capital, the company can raise the capital. However, the “externality” problem will increase the cost of capital. Otherwise, the company can increase the weight of risk-free assets then required capital ratio can be reduced, but the profitability will be limited by this change.

Table 3 shows the result of maximum expected return with the required capital ratio constraint. Given a probability of bankruptcy, increasing the weight of risk-free assets can reduce the required capital ratio. In contrast with Stable (DA) model and Stable (SA) model, the optimal portfolio is quite different in different models. First, the weight of CSCO is much less than the weight of AXP in the Stable (DA) model, but the results are opposite in Stable (SA) model. Second, the weight of risk-free assets in Stable (DA) model is much less than that of Stable (SA) model. This is consistent with Table 2 that reducing the weight of CSCO can rapidly reduce the required capital ratio. In the Stable (SA) model, the allocation of two risky assets is relatively close together, because the extra-ordinary risks of two assets are assumed to be equal in this model.

Table 3 The allocation of AXP and CSCO in Eq. (8) with different capital ratios (cr) and the probability of bankruptcy (θ) is 0.1%

	Model	$r_f = 0.01\%$	AXP	CSCO	Expected Returns	Empirical Results	Min. Returns	The % of bankruptcy
Cr = 30%	Stable (DA)	19.55	65.49	14.96	0.71	0.03	-23.16	0
	Stable (SA)	45.97	16.63	37.4	0.8	0.02	-12.62	0
	Normal	-135.22	118.26	116.96	1.27	0.07	-58.56	1.22
Cr = 20%	Stable (DA)	46.36	43.67	9.98	0.48	0.03	-15.44	0
	Stable (SA)	63.97	11.09	24.94	0.54	0.01	-8.41	0
	Normal	-56.84	78.85	77.99	0.85	0.05	-39.04	1.22
Cr = 10%	Stable (DA)	73.17	21.84	4.99	0.24	0.02	-7.72	0
	Stable (SA)	81.98	5.55	12.48	0.28	0.01	-4.2	0
	Normal	21.54	39.45	39.01	0.43	0.03	-19.53	1.22

Therefore, the company reduced the weight of both assets to decrease the extra-ordinary risk of the total assets in this model, so the weight of risk-free assets is higher in this model than that of the Stable (DA) model. In the Stable (DA) model, the company can keep less risk-free assets by holding less of CSCO, so higher returns are yielded. It is consistent with the empirical results that the returns in the Stable (DA) model are higher than that of the Stable (SA) model without any bankruptcy in both models. In the normal model, although the empirical returns are the highest, but the company gets bankruptcy. It is not preferred in this study (Fig. 2).

85.5.2 US Banks

In tradition, VaR model is used to determine the potential loss of the portfolio and the returns are assumed normally distributed, especially for banks. However, in the crisis period, they find that the potential loss is underestimated by using the VaR model. Therefore, bank regulators proposed the stress test to explain the potential loss in the crisis period. The idea is to assume different scenarios in economic downturns and how they affect the risk-based capital requirement of the banks. However, this method is nonparametric, without involving any probability theory. In our argument, the traditional VaR model cannot capture the extra-ordinary risk, so it underestimates the potential loss in a crisis period. Then, the stress analyzes how the extra-ordinary risk changes the capital requirement of the banks. In our paper, we seem to create a bridge to connect the traditional VaR model and the stress test. Our stable VaR model captures the extra-ordinary risk via the characteristic exponent (α), it gives more accurate estimates of VaR, as shown in previous section. We can also change the

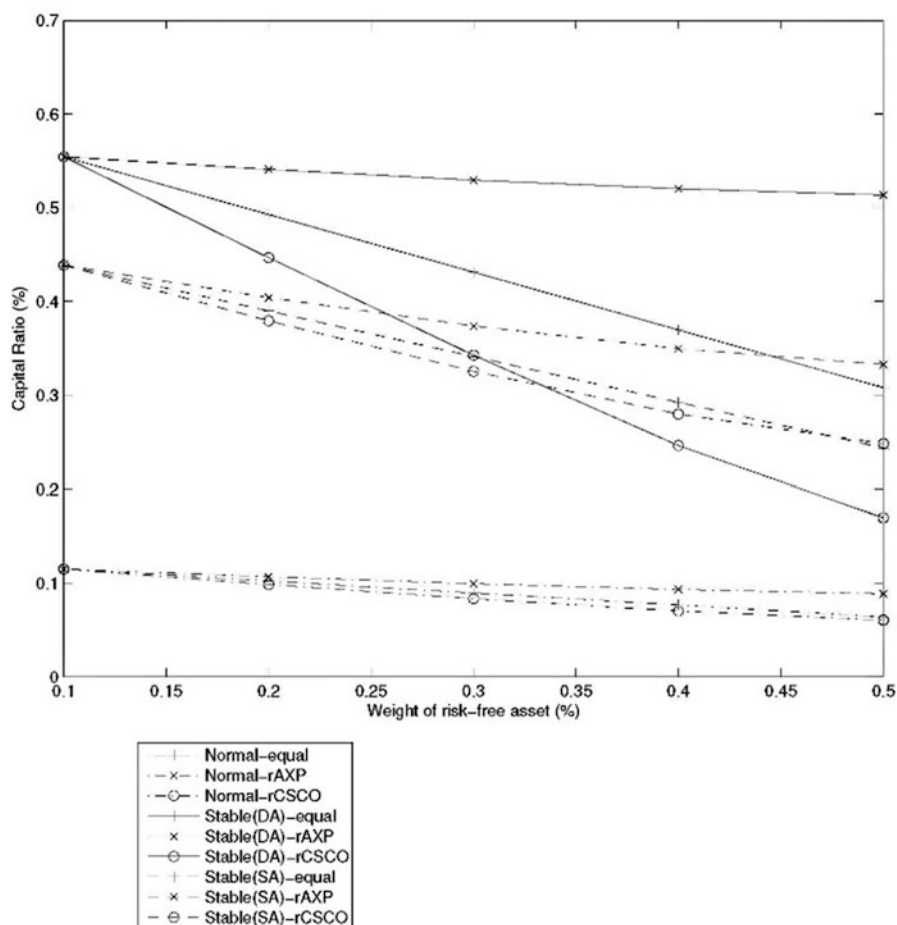


Fig. 1 The description of Table 2

parameters to satisfy different scenarios, for example, we can reduce α to allow the higher extra-ordinary risk in the future, or we may require lower bankruptcy risk (θ). Therefore, the capital ratio will change accordingly. In this chapter, we focus on the required capital ratio with different bankruptcy risks.

The implementation to the Bank of America Corp., the Wells Fargo & Co., the Citigroup Inc., the Regions Financial Corp., and the SunTrust Banks Inc. will be contemplated in this section. For simplicity, the total asset will be classified into four categories, that is, risk-free assets (including cash and balances due from depository institutions, Federal funds sold and securities purchased under agreements to resell), security assets (including all derivatives and securities), loans and leases financing receives (including all loan assets), and other equities (including premises and equipment, goodwill, intangible assets, and other assets). These are obtained from the

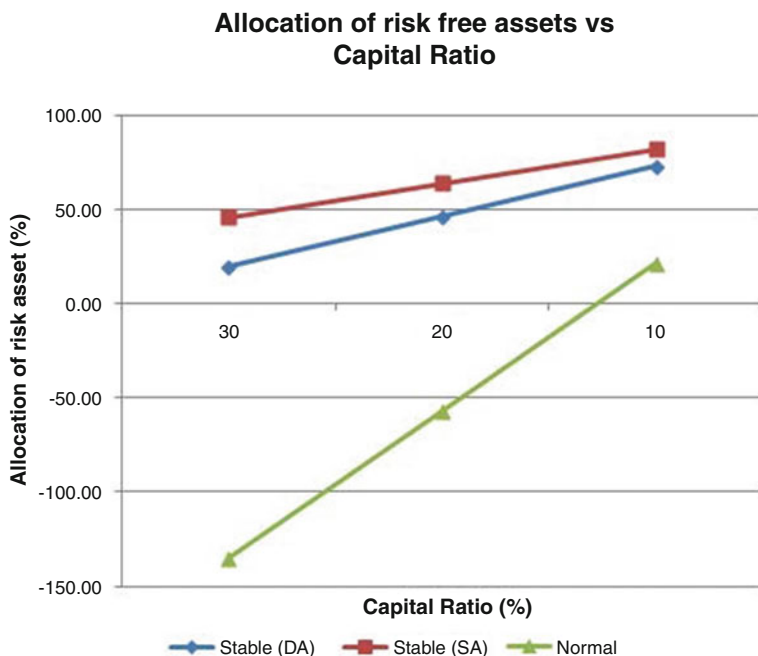


Fig. 2 The description of Table 3

Federal Deposit Insurance Corporation (FDIC) database. The parameters of security assets are estimated by using market index with monthly data from January 1990 to December 2011. Three indices Nikkei 225, S&P 500, and Dax 100 are collected, and the average of these indices is the proxy of the market index in this study. The parameters of loans and leases allowance are proxied by the returns on weighted average of 30-year maturity bond in long position (the proxy of gain from loan) and 1-year maturity bond in short position (the proxy of deposit cost). These indices are collected from the database “CRPS,” then half-year parameters can be evaluated via Eq. (12). The results of monthly parameters are given in Table 4. Other equities are assumed as normal distributed with $c = 0.03$ and $\delta = 0.01$ from 2005 to 2011 (the average value of the assets containing in the other equities). Risk-free rate and interest rate of debt (the average value of deposit, short term, and long-term debts rates) are assumed 1% p.a. and 6% p.a., respectively. We used the consolidated financial statement from the FDIC to find ratios of securities per total asset (S/A), loans and leases per total asset (L/A) and other equities per total asset (O/A). We postulated these assets are statistically independent then we can compute half-year required capital ratios with Eqs. (5), (6) and Theorem 1. The results are given in Tables 7, 8, 9, 10, and 11. The raising capital (RC) is computed through the following equation

$$RC = \frac{rcr \times TA_t - c_t}{1 - rcr}, \quad (13)$$

Table 4 The parameters of bond index (left hand side) and market index (right hand side)

Years	α	β	c	δ	Years	α	β	c	δ
05	1.82	-1	0.018	0.004	05	1.948	-1	0.031	0.006
06	1.784	-1	0.017	0.004	06	1.943	-1	0.03	0.007
07	1.799	-1	0.017	0.004	07	1.987	-1	0.029	0.007
08	1.873	-0.84	0.019	0.006	08	1.791	-1	0.029	0.007
09	2	0	0.02	0.005	09	1.772	-1	0.031	0.007
10	1.961	-1	0.02	0.005	10	1.852	-1	0.032	0.01
11	1.935	-1	0.02	0.005	11	1.801	-1	0.031	0.006

Table 5 The details of Treasury's bank bailout between 2008 and 2011, the details can be found at <http://money.cnn.com/news/specials/storysupplement/bankbailout/>

Name of banks	Treasury's bank bailout
Bank of America	\$15 billion in October 28, 2008 ^a and addition \$10 billion in January 9, 2009.
Wells Fargo	\$25 billion in October 28, 2008.
Citigroup	\$25 billion in October 28, 2008.
Regions financial	\$3.5 billion in November 17, 2008.
SunTrust	\$3.5 billion in November 17, 2008, \$1.35 billion in December 31, 2008.

^aInitially allocated to Merrill Lynch, which Bank of America purchased on Jan. 1, 2009 Source: Treasury Dept.

where c_t is the capital at time t and rcr is the required capital ratio with probability of default θ in Eq. (7).

As shown in Tables 7, 8, 9, 10, and 11, all banks need to raise their capital before 2008. For example, the Bank of America needs more than 12% capital ratio in 2007 such that it can reduce the probability of default to 1.5%, it needs to increase more than \$50 billion in capital. This was a warning of the US banking system. After that, a financial crisis started in 2008 and the treasury department injected billions of dollars into the financial market. The amounts are similar to our estimations, that is, the Wells Fargo was injected with \$25 billion that is close to our estimation of \$23.06 billion in the fourth quarter of 2008 with 4% probability of default (see Table 8). For Citigroup, the amount was \$25 billion that is similar to our estimation of \$20.97 billion (see Table 9). For Regions Financial, the amount is \$3.5 billion that is close to our estimation of \$3.22 billion (see Table 10). On the other hand, three large banks (the Bank of America, the Wells Fargo, and the Citigroup) have lower capital ratios than others, and all of them are less than 10% before 2008. The probability of default was more than 4% before 2008. That is a dangerous signal for the banking system.

First, all banks in our study got the bailout from the US government (see Table 5) and the capital ratio increased after the Treasury's bank bailout. The government also wants to stimulate the economy. The amount of loans should be increased significantly. However, all banks reduced the L/A ratio and increased the weight of risk-free assets from the first quarter of 2009 to the end of 2009. On the other hand, if we postulated the worst economy in the future, banks may raise their capital until it

reaches to the required capital ratio with a lower probability of bankruptcy, such as 1%. It can be viewed as a stress test with different scenarios, that is, the Bank of America needed to raise its capital ratio to 11.62% with 1.5% probability of default or the bank may need to raise the capital ratio to 13.53% with 1% probability of default. The required capital becomes much more than that needed at 1.5%. However, it will cause the bank loans to decrease and the cost of capital to increase, and it may slow down the economic growth in USA. The government should recover the confidence of investors but minimize the impact of the weak economic environment. It can be concluded as reducing the bankruptcy risk with minimal rising of capital. In this chapter, we recommend holding a suitable weight of risk-free assets instead of raising capital. It can reduce the probability of bankruptcy and lighten the increase in cost of capital. As shown in Tables 7, 8, and 9, the Bank of America increased the weight of risk-free assets from 6.76% in 2008 to 16.01% in 2010; the Wells Fargo increased from 5.59% in 2008 to 7.61% in 2010; the Citigroup increased from 19.79% in 2008 to 22.82% in 2010. Figs. 3 and 4 show the changes in the weight of risk-free assets amongst different banks. In Fig. 3, the total amount of cash increased significantly in the first quarter of 2009, especially for the Bank of America and the Citigroup. In Fig. 4, the “Federal funds sold and securities purchased under agreements to resell” of the Bank of America are increased substantially after 2008, and the Citigroup is at a peak before 2008, because it suffered a lot in subprime mortgage crisis since it invested a huge amount in Collateralized Debt Obligations (CDO) before the crisis. If the Federal funds sold and securities purchased under agreements to resell are viewed as US government subsidy through the Quantitative Easing (QE).¹⁴ Fig. 5 depicts the amount of loan decreases after 2008. It shows that QE cannot increase the amount of loan after the US bank regulators’ announcement. They increase the weight of risk-free assets such that they can reach the capital requirement. It explains why banks do not increase the loan significantly. The small banks (the Regions Financial & the SunTrust) do not increase their weight of risk-free assets, since the raising capital is around \$2.5 billion, which is relatively smaller than the big banks (Fig. 6).

Second, we estimate the parameters of the different banks by using their end-of-day stock prices (the average of bid-ask prices) from 2003 to 2011. As shown in Table 6, the extraordinary risks of the banks are at the peak in the crisis period. The stock prices slumped more than 30%, thus capital ratios reduced before government bailout. After 2008, the banks got the bailout and reallocated their portfolios. One of the reasons is to achieve the capital requirement, and another reason is to stay away from the market. Two reasons are correlated, once the money is drawn from the market, the required capital ratio reduces, and the banks can reach the capital requirement without raising billions of dollars in their capital. After the crisis, the extra-ordinary risk of the market decreases, and the bankruptcy risks of the banks

¹⁴The QE1 is announced in November 2008.

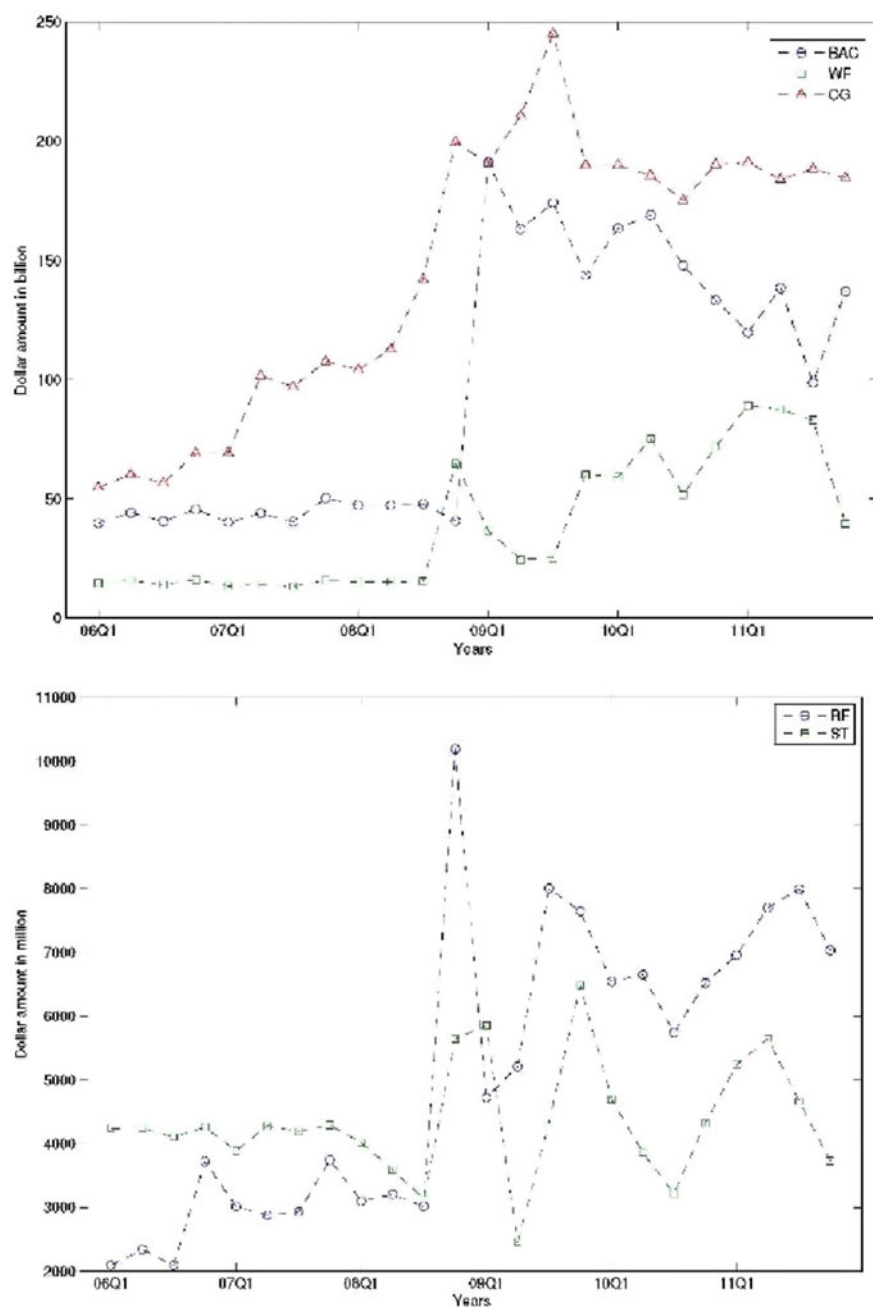


Fig. 3 The description of cash in banks

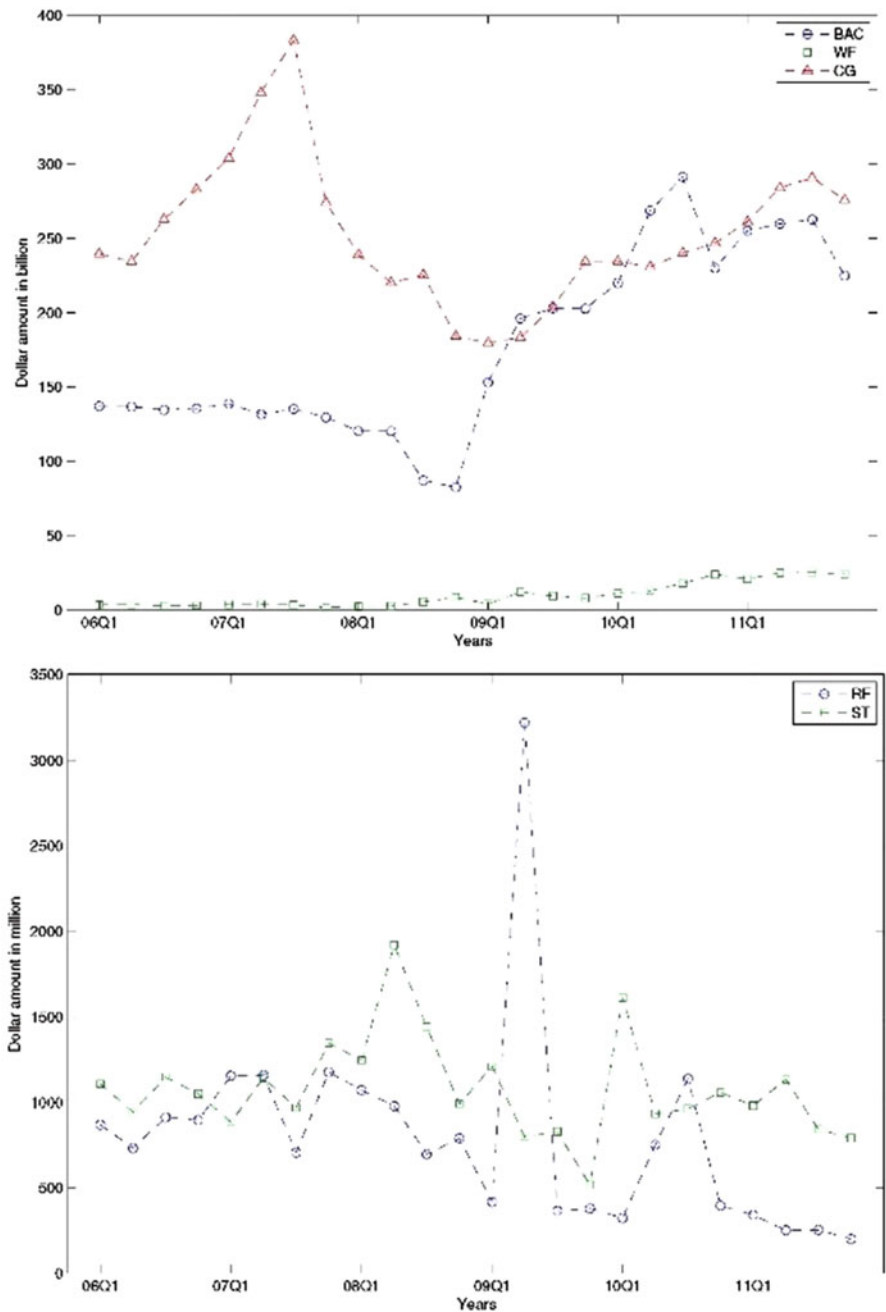


Fig. 4 The description of Federal funds sold and securities purchased under agreements to resell in banks

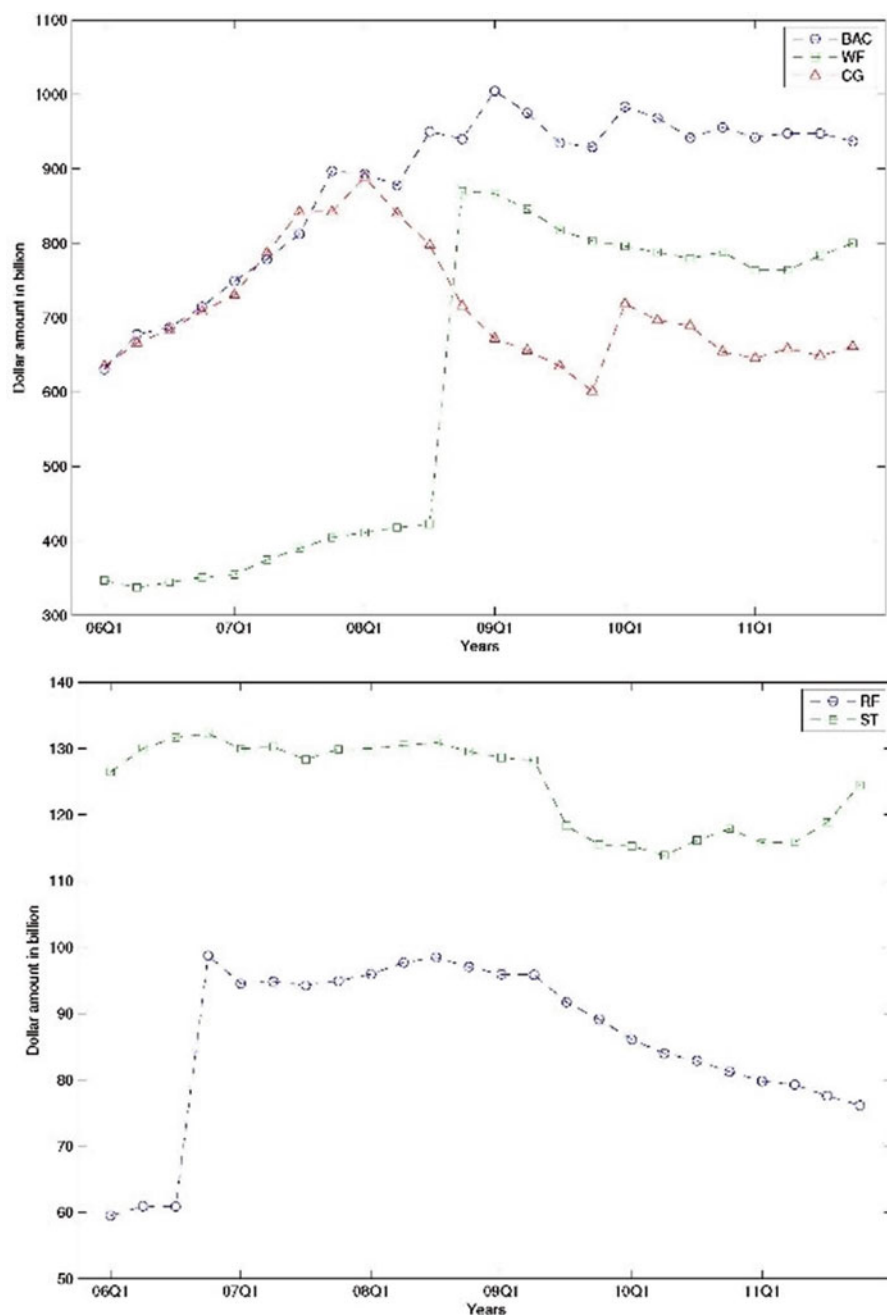


Fig. 5 The description of loans and leases financially received in Banks

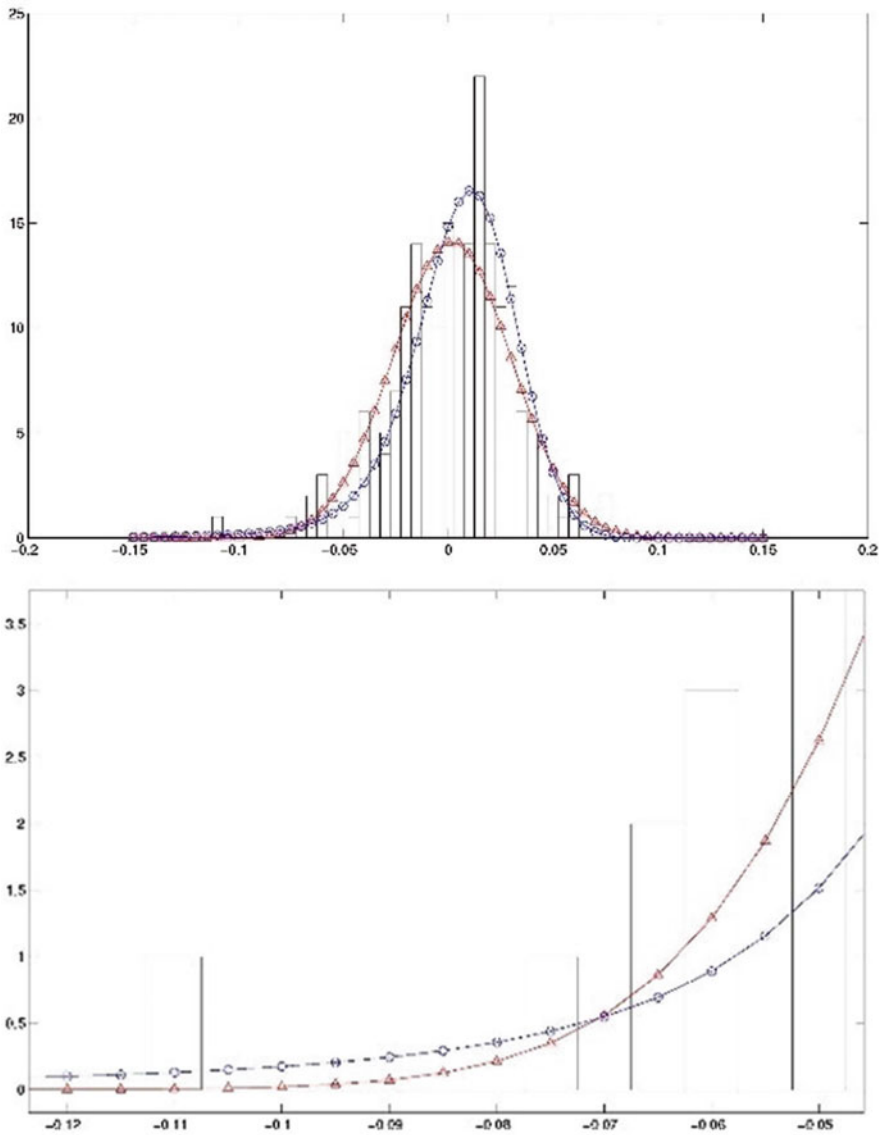


Fig. 6 The density functions of normal distribution with $\alpha = 2, \beta = 0, c = 0.02$ and $\delta = 0.002$ (the line with triangles), stable distribution with $\alpha = 1.784, \beta = -1, c = 0.017$ and $\delta = 0.004$ (the line with circles) and the empirical distribution (the histogram) of weekly bond returns in 2006

reduce. Therefore, their capital is more sufficient than that in the crisis period. On January 4, 2013, “Big Banks Readying Payout Plans,” *The Wall Street Journal*,

Citigroup will seek permission for its first share buyback since 2007 as part of the latest Federal Reserve “stress tests,” according to people familiar with the company’s plans.

Table 6 The monthly parameters of returns of different banks. First, the daily returns are used to estimate the parameters, and then Eq. (12) is used to aggregate the daily parameters to monthly parameters

	Before the crisis (2004 to 2006)				At the crisis (2007 to 2010)				After the crisis (2011)			
	α	β	c	δ	α	β	c	δ	α	β	c	δ
BAC	1.859	-0.357	0.03	0.017	1.257	0	0.177	-0.024	1.491	0	0.129	-0.073
WF	1.877	0.087	0.028	0.006	1.305	0	0.158	-0.023	1.765	0	0.081	-0.007
CG	1.772	0.084	0.036	0.006	1.299	0	0.186	-0.038	1.469	-0.082	0.129	-0.025
RF	1.701	0.100	0.042	0.001	1.364	0	0.199	-0.032	1.647	-0.262	0.128	-0.002
ST	1.800	0.389	0.033	-0.004	1.35	0	0.179	-0.024	1.699	0	0.104	-0.035

“These banks are strong and have too much capital now,” said RBC Capital Market’s analyst Gerard Cassidy, who predicts all major banks that ask for dividends and buybacks in 2013 will be approved.

The banks have surplus capital, because the extraordinary risk of the market is reduced and the portfolio reallocated. They require dividend payout or share buybacks. In Miller and Rock (1985), they introduce information signal model and conclude that the dividend policy gives signals of future prospects of companies. Increasing the dividend payout signals good news and share prices increase. Braggion and Moore (2011) find the abnormal returns when dividend policy is changed. Therefore, the banks ask for the dividend payout to push up their stock prices, so that the capital ratio increased. The banks can reach the capital requirement.

Third, we observe the allowance of bankruptcy risks of different banks are different. In the Bank of America, it is required to raise \$33.9 billion. The close estimation is \$33.93 billion with 1.5% bankruptcy risk. In the Wells Fargo and the Citigroup, the closest values of raising capital are \$14.6 billion and \$4.39 billion, respectively, which are in the 4% probability of default. The Regions Financial is \$2.93 billion with 1.25% bankruptcy risk and the SunTrust is \$2.58 billion with 1.5% bankruptcy risk. Our explanations are (1) the assets of the Bank of America, the Citigroup, and the Wells Fargo are almost ten times the other two banks. The US government may allow a higher probability of bankruptcy for banks with huge amounts of assets because they can change their portfolio easier than small banks. (2) The structure of loans of different banks may not be the same, while we aggregate all of them (It may include mortgage loans and commercial loans, etc.) as one component. It may cause the difference. (3) The numerical error of our algorithm, we estimate required capital ratios, but the raising capital is very sensitive to the required capital ratio. Therefore, the error in the raising capital enlarges. The Bank of America requires the lowest probability of bankruptcy (1.5%) than other big banks (the Wells Fargo and the Citigroup). One main reason is the big loss suffered by acquiring the Merrill Lynch in December 2008. Merrill Lynch reported the loss of \$15.3 billion in the fourth quarter of 2008 and the Bank of America suffered the loss of \$ 1.75 billion, the stock price is down at \$7.18, dropped by 13.7%. Mr. Lewis (the chief executive of the Bank of America) said, “. . . He was surprised to learn in December, three months after the bank snapped up Merrill Lynch in a shotgun deal, that the magnitude of losses at the

brokerage was far greater than expected. He said he had considered walking away from the deal at that point, but was persuaded not to, partly by regulators who feared that a failure to seal the deal could set off a new round of panic in the markets. . .” (See “For Bank of America, the Pressure Mounts Over Merrill Deal” by LOUISE STORY in January 17, 2009, in *The New York Times*).

Fourth, the Wells Fargo and the Citigroup are trying to increase their capital into a more safe level. They kept increasing their capital ratios from 2009 to 2011, therefore the bankruptcy risk was reduced from 4% to 3%. However, the Sun Trust increased the weight of security assets from 13.93% in the first quarter of 2009 to 17.27% at the end of 2011 and reduced the weight of risk-free assets to 2.56%. Therefore, its bankruptcy risk increases from 1.5% to 2.5%. Regions Financial also increased the weight of security assets from 15.54% in the first quarter of 2009 to 20.05% at the end of 2011, but it also increased the weight of risk-free assets from 3.62% to 5.69%. Therefore, its bankruptcy risk remains at 1.25%. The bankruptcy risk of the Bank of America is around 2% in 2011. It may not be worse, since the subprime mortgage crisis is over, and the big loss from the Merrill Lynch has subsided. The government may allow a higher bankruptcy risk since the Bank of America is one of the Big Four banks in America. Big banks tend to reduce the bankruptcy risk through increasing the risk-free weight and the capital, but small banks tend to increase their profits through increasing the weight of security assets and decreasing loans and leases.

85.6 Conclusion

This study proposes a VaR and stress test-integrated method. We introduce a VaR-based method to evaluate the required capital ratio. After that, the comparison of VaR estimations in different models is proposed. In June 2009, J.P. Morgan Investment Analytics and Consulting Romain Berry describe “. . . a risk manager should not rely too much on VaR calculations since they may be underestimated quite often. . .” This chapter tackles this issue by using the stable Paretian distribution instead of the normal distribution and statistical evidence is given. The normal distribution captures all risks in one parameter. On the other hand, stable Paretian distribution separates the risks into extraordinary risk and ordinary risk with different parameters. It is more suitable for the risk managers to make decisions in normal period and also the crisis period. Some authors apply the extreme value theory to the VaR. However, it may not be suitable in the normal period and are not flexibly used to maximize expected returns, since the idea of extreme value theory only considers the tail of distributions. As shown in Fig. 7, the normal distribution tries to capture the extreme returns by increasing the standard derivation (σ). However, it will underestimate the probabilities around the mean. In stable Paretian distribution, the extra-ordinary risk and ordinary risk are captured by different parameters α and c . Therefore, it fits the empirical distribution much better. A Fourier series-based method gives us a new tool to calculate the stable Paretian distribution with different α . Therefore, by separating the extra-ordinary risk of assets, company may find a better asset structure such that higher profit can be obtained. The empirical results in Tables 2 and 3 support these arguments. Finally, we find out five banks that are required to raise the capital. As shown in Tables 7, 8, 9, 10, and 11, all

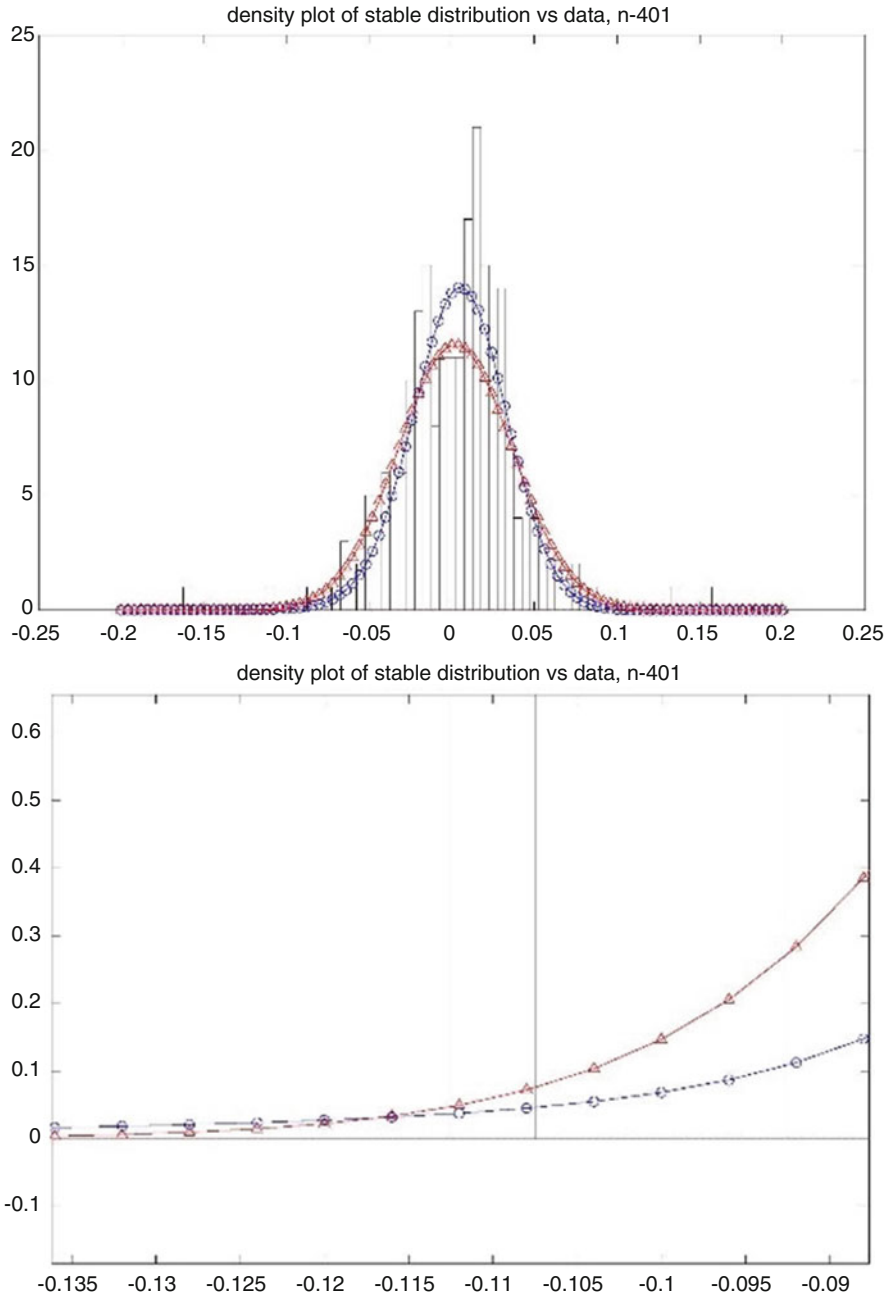


Fig. 7 The density functions of normal distribution with $\alpha = 2, \beta = 0, c = 0.024$ and $\delta = 0.002$ (the line with triangles), stable distribution with $\alpha = 1.961, \beta = -1, c = 0.02$ and $\delta = 0.005$ (the line with circles) and the empirical distribution (the histogram) of weekly bond returns in 2010

Table 7 The estimations of required capital ratios of the Bank of America, all units are in percentage except raising capital in billion. w_0 is the weight of risk-free assets, the column cr is capital ratios in the balance sheet of the bank. Columns 3%, 2.5%, 2%, 1.5%, and 1% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	cr	3%	2.50%	2%	1.50%	1%
06Q1	12.85	24.02	45.82	17.31	9.41	9.37 (0)	9.91 (7.63)	10.6 (18.42)	11.58 (33.89)	13.2 (60.15)
06Q2	12.49	23.4	46.81	17.3	8.83	9.4 (9.16)	9.95 (18.02)	10.66 (29.63)	11.66 (46.3)	13.3 (74.67)
06Q3	12.06	20.92	47.28	19.73	9.2	9.26 (0.97)	9.81 (9.74)	10.51 (21.27)	11.51 (37.89)	13.17 (66.28)
06Q4	12.37	21.14	48.81	17.68	9.24	9.56 (5.1)	10.18 (15.37)	11.01 (29.09)	12.2 (49.3)	14.2 (84.62)
07Q1	11.84	20.91	49.62	17.63	8.93	9.63 (11.65)	10.27 (22.43)	11.1 (36.87)	12.31 (58.14)	14.34 (95.36)
07Q2	11.43	20.92	50.7	16.95	8.84	9.75 (15.43)	10.4 (26.66)	11.25 (41.7)	12.48 (63.88)	14.55 (102.71)
07Q3	11.08	20.34	51.35	17.23	8.75	9.78 (18.07)	10.44 (29.77)	11.3 (45.46)	12.55 (68.62)	14.64 (109.2)
07Q4	10.45	21.71	52.13	15.71	8.53	9.54 (19.18)	10.15 (30.92)	10.94 (46.52)	12.08 (69.35)	13.98 (109.06)
08Q1	9.61	24.78	51.23	14.38	8.97	9.59 (12.14)	10.19 (23.79)	10.97 (39.2)	12.07 (61.6)	13.92 (100.37)
08Q2	8.79	26.95	50.94	13.31	9.44	9.69 (4.78)	10.28 (16.2)	11.05 (31.25)	12.15 (53.05)	13.96 (90.59)
08Q3	7.35	26.29	51.73	14.63	8.77	9.74 (19.69)	10.34 (32.14)	11.12 (48.57)	12.23 (72.41)	14.08 (113.57)
08Q4	6.76	25.6	51.59	16.05	9.72	9.95 (4.68)	10.63 (18.66)	11.53 (37.24)	12.8 (64.42)	14.93 (111.58)
09Q1	14.82	23.48	43.24	18.46	10.33	9.07 (0)	9.68 (0)	10.48 (3.86)	11.62 (33.93)	13.53 (86.04)
09Q2	15.92	22.4	43.22	18.47	11.33	8.92 (0)	9.52 (0)	10.29 (0)	11.4 (1.87)	13.26 (50.22)
09Q3	16.74	22.85	41.49	18.93	11.46	8.87 (0)	9.46 (0)	10.23 (0)	11.34 (0)	13.2 (45.05)
09Q4	15.53	23.93	41.65	18.9	10.41	9.08 (0)	9.62 (0)	10.31 (0)	11.28 (21.85)	12.9 (63.86)
10Q1	16.34	23.83	41.9	17.93	9.87	9.07 (0)	9.6 (0)	10.29 (11.1)	11.26 (36.85)	12.87 (81.03)
10Q2	18.46	23.63	40.84	17.07	9.91	8.98 (0)	9.51 (0)	10.19 (7.25)	11.15 (33.06)	12.76 (77.44)
10Q3	18.76	25.4	40.23	15.61	9.92	9.26 (0)	9.83 (0)	10.56 (16.86)	11.61 (44.76)	13.36 (93.17)
10Q4	16.01	25.51	42.12	16.36	10.46	8.56 (0)	9.04 (0)	9.66 (0)	10.51 (1.27)	11.85 (35.88)
11Q1	16.46	25.12	41.39	17.02	10.19	8.47 (0)	8.95 (0)	9.56 (0)	10.39 (5.02)	11.71 (39.25)
11Q2	17.58	24.63	41.85	15.94	9.86	8.42 (0)	8.9 (0)	9.5 (0)	10.33 (11.68)	11.63 (45.29)
11Q3	16.26	25.62	42.65	15.47	10.39	8.6 (0)	9.09 (0)	9.71 (0)	10.56 (4.14)	11.91 (38.27)
11Q4	16.93	24.6	43.85	14.62	10.8	9.58 (0)	10.16 (0)	10.91 (2.72)	11.97 (28.41)	13.72 (72.27)

Table 8 The estimations of required capital ratios of the Wells Fargo, all units are in percentage except raising capital is in billion. w_0 is the weight of risk-free assets, the column cr is capital ratios in the balance sheet of the bank. Columns 4.5%, 4%, 3.5%, 3%, and 2.5% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	cr	4.50%	4%	3.50%	3%	2.50%
06Q1	3.69	11.49	70.46	14.36	8.52	9.84 (7.23)	10.28 (9.67)	10.79 (12.54)	11.4 (16.01)	12.16 (20.38)
06Q2	3.89	15.26	67.48	13.36	8.39	9.64 (6.91)	10.06 (9.3)	10.56 (12.11)	11.15 (15.51)	11.88 (19.78)
06Q3	3.45	11.4	71.2	13.94	9.28	9.92 (3.41)	10.36 (5.81)	10.87 (8.64)	11.49 (12.05)	12.25 (16.36)
06Q4	3.96	9.41	72.86	13.77	9.52	10.23 (3.83)	10.73 (6.56)	11.32 (9.79)	12.03 (13.75)	12.92 (18.82)
07Q1	3.52	10.07	73.03	13.38	9.49	10.25 (4.08)	10.75 (6.84)	11.34 (10.12)	12.05 (14.13)	12.94 (19.24)
07Q2	3.3	14.03	69.36	13.31	8.76	9.92 (6.96)	10.4 (9.9)	10.97 (13.39)	11.65 (17.65)	12.5 (23.08)
07Q3	3.04	11.27	71.04	14.65	8.7	10.06 (8.28)	10.55 (11.33)	11.12 (14.96)	11.82 (19.39)	12.69 (25.05)
07Q4	3.04	13.53	70.32	13.12	8.28	9.81 (9.78)	10.27 (12.8)	10.81 (16.36)	11.46 (20.69)	12.27 (26.19)
08Q1	2.91	14.97	69.06	13.06	8.09	9.7 (10.58)	10.15 (13.66)	10.68 (17.28)	11.32 (21.68)	12.12 (27.26)
08Q2	2.9	16.26	68.59	12.24	7.87	9.67 (12.08)	10.12 (15.21)	10.65 (18.89)	11.28 (23.37)	12.07 (29.05)
08Q3	3.37	15.12	67.9	13.62	7.54	9.58 (14.04)	10.03 (17.21)	10.56 (20.95)	11.18 (25.5)	11.97 (31.25)
08Q4	5.59	15.5	66.44	12.48	7.57	8.77 (17.31)	9.16 (23.06)	9.62 (29.73)	10.15 (37.7)	10.81 (47.58)
09Q1	3.17	17.29	67.38	12.16	8.33	8.95 (8.79)	9.35 (14.6)	9.82 (21.35)	10.38 (29.42)	11.05 (39.45)
09Q2	2.85	19.02	65.88	12.26	9.45	8.94 (0)	9.35 (0)	9.83 (5.33)	10.39 (13.37)	11.07 (23.38)
09Q3	2.83	18.24	66.52	12.42	10.49	8.94 (0)	9.35 (0)	9.82 (0)	10.38 (0)	11.06 (7.8)
09Q4	5.46	17.08	64.57	12.88	9.2	8.64 (0)	8.94 (0)	9.28 (1.18)	9.67 (6.5)	10.12 (12.81)
10Q1	5.76	16.81	65.07	12.36	9.66	8.66 (0)	8.96 (0)	9.3 (0)	9.68 (0.32)	10.13 (6.47)
10Q2	7.18	16.47	64.25	12.09	9.9	8.57 (0)	8.87 (0)	9.2 (0)	9.58 (0)	10.02 (1.64)
10Q3	5.72	18.25	63.87	12.17	10.25	8.68 (0)	8.99 (0)	9.33 (0)	9.73 (0)	10.2 (0)
10Q4	7.61	17.43	62.66	12.31	10.17	8.28 (0)	8.59 (0)	8.93 (0)	9.33 (0)	9.81 (0)
11Q1	8.84	17.73	61.36	12.07	10.84	8.19 (0)	8.5 (0)	8.84 (0)	9.24 (0)	9.71 (0)
11Q2	8.93	18.71	60.63	11.73	10.95	8.18 (0)	8.49 (0)	8.84 (0)	9.24 (0)	9.72 (0)
11Q3	8.29	19.95	60.04	11.72	10.67	8.21 (0)	8.52 (0)	8.88 (0)	9.28 (0)	9.77 (0)
11Q4	4.86	22.53	60.89	11.73	10.78	9.25 (0)	9.62 (0)	10.05 (0)	10.56 (0)	11.17 (5.65)

Table 9 The estimations of required capital ratios of the Citigroup, all units are in percentage except raising capital is in billion. w_0 is the weight of risk-free assets, the column cr is capital ratios in the balance sheet of the bank. Columns 4.5%, 4%, 3.5%, 3%, and 2.5% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	cr	4.50%	4%	3.50%	3%	2.50%
06Q1	18.55	22.69	40.02	18.75	7.21	7.68 (8)	7.97 (13.07)	8.31 (18.93)	8.7 (25.89)	9.19 (34.46)
06Q2	18.11	21.7	40.95	19.24	7.1	7.67 (10.11)	7.96 (15.34)	8.3 (21.4)	8.7 (28.6)	9.19 (37.5)
06Q3	18.27	23.83	39.15	18.75	6.75	7.72 (18.31)	8.01 (23.91)	8.34 (30.38)	8.74 (38.05)	9.22 (47.48)
06Q4	18.67	24.64	37.59	19.09	6.36	7.63 (26.05)	7.94 (32.41)	8.29 (39.81)	8.71 (48.64)	9.23 (59.61)
07Q1	18.47	24.55	36.13	20.84	6.04	7.54 (32.68)	7.83 (39.34)	8.18 (47.06)	8.59 (56.27)	9.09 (67.68)
07Q2	20.25	21.37	35.44	22.95	5.75	7.23 (35.48)	7.52 (42.44)	7.85 (50.52)	8.24 (60.18)	8.72 (72.18)
07Q3	20.36	21.9	35.7	22.04	5.39	7.29 (48.26)	7.58 (55.76)	7.91 (64.48)	8.3 (74.89)	8.79 (87.82)
07Q4	17.45	20.19	38.5	23.86	5.19	7.11 (45.24)	7.39 (51.84)	7.7 (59.48)	8.07 (68.59)	8.53 (79.85)
08Q1	15.6	27.3	40.4	16.7	5.83	7.66 (43.75)	7.96 (50.96)	8.3 (59.27)	8.7 (69.12)	9.18 (81.2)
08Q2	15.85	27.14	40	17.01	6.49	7.63 (25.72)	7.92 (32.5)	8.25 (40.31)	8.65 (49.55)	9.13 (60.89)
08Q3	17.92	25.7	38.89	17.5	6.15	7.44 (28.6)	7.72 (35)	8.05 (42.39)	8.43 (51.14)	8.9 (61.89)
08Q4	19.79	27.37	36.91	15.93	7.31	7.93 (13.08)	8.3 (20.97)	8.73 (30.22)	9.24 (41.41)	9.89 (55.5)
09Q1	20.31	26.88	36.85	15.96	8.01	7.86 (0)	8.23 (4.39)	8.65 (12.9)	9.16 (23.19)	9.8 (36.15)
09Q2	21.27	27.45	35.42	15.85	8.32	7.87 (0)	8.24 (0)	8.66 (6.88)	9.18 (17.35)	9.81 (30.55)
09Q3	23.71	27.2	33.54	15.55	7.55	7.75 (4.12)	8.11 (11.55)	8.53 (20.28)	9.03 (30.84)	9.66 (44.16)
09Q4	22.32	29.88	32.28	15.51	8.32	8.49 (3.48)	8.88 (11.36)	9.33 (20.62)	9.86 (31.84)	10.53 (46.01)
10Q1	21.19	29.16	35.89	13.76	7.68	8.48 (17.42)	8.85 (25.77)	9.29 (35.56)	9.81 (47.37)	10.46 (62.24)
10Q2	21.48	28.84	35.95	13.72	8.12	8.43 (6.51)	8.8 (14.46)	9.23 (23.77)	9.75 (35.01)	10.39 (49.16)
10Q3	20.93	31.24	34.74	13.09	8.33	8.78 (9.72)	9.18 (18.51)	9.64 (28.86)	10.2 (41.39)	10.9 (57.21)
10Q4	22.82	29.82	34.17	13.19	8.66	7.59 (0)	7.92 (0)	8.29 (0)	8.73 (1.34)	9.26 (12.56)
11Q1	23.22	29.85	33.14	13.79	8.9	7.56 (0)	7.88 (0)	8.25 (0)	8.69 (0)	9.22 (6.8)
11Q2	23.91	28.76	33.64	13.68	9.13	7.44 (0)	7.76 (0)	8.12 (0)	8.54 (0)	9.06 (0)
11Q3	24.74	28.85	33.5	12.9	9.26	7.45 (0)	7.76 (0)	8.12 (0)	8.55 (0)	9.06 (0)
11Q4	24.57	27.58	35.27	12.57	9.58	8.32 (0)	8.68 (0)	9.1 (0)	9.59 (0.17)	10.2 (12.9)

Table 10 The estimations of required capital ratios of the Regions Financial, all units are in percentage except raising capital is in billion. w_0 is the weight of risk-free assets, the column cr is capital ratios in the balance sheet of the bank. Columns 3%, 2.5%, 2%, 1.5%, and 1% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	cr	3%	2.5%	2%	1.5%	1%
06Q1	3.51	14.43	70.33	11.73	12.6	11.45 (0)	12.2 (0)	13.2 (0.59)	15.63 (3.04)	16.99 (4.48)
06Q2	3.57	14.33	70.78	11.31	12.43	11.5 (0)	12.26 (0)	13.26 (0.82)	15.7 (3.34)	17.08 (4.82)
06Q3	3.46	15.44	69.98	11.12	12.7	11.44 (0)	12.19 (0)	13.18 (0.49)	15.6 (2.99)	16.96 (4.47)
06Q4	3.22	13.4	68.86	14.52	14.44	11.58 (0)	12.42 (0)	13.55 (0)	16.33 (3.23)	17.9 (6.04)
07Q1	3.02	13.86	68.45	14.67	14.71	11.54 (0)	12.38 (0)	13.5 (0)	16.26 (2.55)	17.82 (5.22)
07Q2	2.93	13.24	68.89	14.94	14.31	11.58 (0)	12.43 (0)	13.55 (0)	16.33 (3.32)	17.9 (6.02)
07Q3	2.63	12.75	68.19	16.42	14.36	11.49 (0)	12.33 (0)	13.44 (0)	16.19 (3.03)	17.75 (5.7)
07Q4	3.49	12.71	67.29	16.51	14.05	11.1 (0)	11.88 (0)	12.9 (0)	15.44 (2.31)	16.87 (4.78)
08Q1	2.89	13.74	66.51	16.86	13.88	11.01 (0)	11.78 (0)	12.79 (0)	15.3 (2.42)	16.72 (4.91)
08Q2	2.89	12.75	67.65	16.7	13.64	11.14 (0)	11.92 (0)	12.96 (0)	15.5 (3.18)	16.94 (5.74)
08Q3	2.57	12.4	68.23	16.8	13.66	11.21 (0)	12 (0)	13.04 (0)	15.61 (3.34)	17.06 (5.92)
08Q4	7.51	13.34	66.35	12.81	11.5	10 (0)	10.63 (0)	11.45 (0)	13.4 (3.22)	14.48 (5.11)
09Q1	3.62	15.54	67.55	13.29	12.04	10.26 (0)	10.92 (0)	11.78 (0)	13.82 (2.93)	14.96 (4.87)
09Q2	5.9	14.24	67.11	12.74	13.32	10.13 (0)	10.78 (0)	11.61 (0)	13.61 (0.48)	14.71 (2.34)
09Q3	5.97	15.61	65.43	13	13.39	10.07 (0)	10.71 (0)	11.55 (0)	13.56 (0.26)	14.67 (2.1)
09Q4	5.64	18.55	62.61	13.2	12.76	9.67 (0)	10.13 (0)	10.71 (0)	11.97 (0)	12.62 (0)
10Q1	5	18.09	62.7	14.2	13.04	9.63 (0)	10.09 (0)	10.66 (0)	11.9 (0)	12.53 (0)
10Q2	5.47	18.4	62.02	14.11	13.09	9.61 (0)	10.07 (0)	10.64 (0)	11.9 (0)	12.54 (0)
10Q3	5.15	18.64	62.06	14.15	13.05	9.64 (0)	10.1 (0)	10.68 (0)	11.95 (0)	12.6 (0)
10Q4	5.22	18.06	61.35	15.36	12.83	9.27 (0)	9.74 (0)	10.33 (0)	11.61 (0)	12.27 (0)
11Q1	5.54	19.31	60.55	14.61	12.81	9.29 (0)	9.77 (0)	10.37 (0)	11.68 (0)	12.35 (0)
11Q2	6.07	18.8	60.55	14.58	13.1	9.25 (0)	9.73 (0)	10.32 (0)	11.61 (0)	12.28 (0)
11Q3	6.36	19.79	59.78	14.07	13.5	9.26 (0)	9.74 (0)	10.34 (0)	11.66 (0)	12.34 (0)
11Q4	5.69	20.05	59.93	14.33	13.19	10.19 (0)	10.76 (0)	11.49 (0)	13.18 (0)	14.09 (1.34)

Table 11 The estimations of required capital ratios of the SunTrust, all units are in percentage except raising capital is in billion. w_0 is the weight of risk-free assets, the column cr is capital ratios in the balance sheet of the bank. Columns 3%, 2.5%, 2%, 1.5%, and 1% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	cr	3%	2.5%	2%	1.5%	1%
06Q1	2.99	16.2	70.69	10.12	9.59	11.54 (3.95)	12.31 (5.53)	13.3 (7.66)	14.73 (10.79)	17.12 (16.24)
06Q2	2.86	15.65	71.77	9.72	9.62	11.65 (4.16)	12.42 (5.8)	13.43 (7.98)	14.88 (11.2)	17.3 (16.82)
06Q3	2.87	15.49	71.9	9.74	10.15	11.66 (3.12)	12.43 (4.77)	13.45 (6.97)	14.9 (10.21)	17.32 (15.87)
06Q4	2.91	14.74	72.57	9.78	9.78	12.05 (4.71)	12.94 (6.62)	14.12 (9.21)	15.82 (13.08)	18.67 (19.92)
07Q1	2.56	17.42	69.72	10.3	9.64	11.78 (4.52)	12.64 (6.39)	13.77 (8.93)	15.41 (12.72)	18.17 (19.43)
07Q2	3.01	14.07	72.21	10.71	9.63	12 (4.85)	12.88 (6.72)	14.05 (9.27)	15.75 (13.09)	18.59 (19.84)
07Q3	2.93	12.73	72.97	11.36	10.18	12.07 (3.77)	12.96 (5.61)	14.14 (8.11)	15.85 (11.85)	18.72 (18.47)
07Q4	3.14	14.09	72.33	10.44	10.05	11.71 (3.36)	12.54 (5.1)	13.63 (7.44)	15.21 (10.93)	17.87 (17.08)
08Q1	2.94	13.57	72.61	10.88	10.3	11.74 (2.92)	12.57 (4.65)	13.67 (6.99)	15.26 (10.48)	17.92 (16.62)
08Q2	3.11	12.79	73.61	10.49	10.1	11.85 (3.52)	12.7 (5.27)	13.82 (7.63)	15.43 (11.15)	18.12 (17.36)
08Q3	2.62	11.84	74.88	10.66	10.27	12.01 (3.44)	12.87 (5.2)	14 (7.57)	15.64 (11.11)	18.38 (17.35)
08Q4	3.51	14.43	68.49	13.58	11.84	10.27 (0)	10.93 (0)	11.78 (0)	12.98 (2.48)	14.94 (6.89)
09Q1	3.94	13.93	71.76	10.37	12.08	10.55 (0)	11.23 (0)	12.1 (0.05)	13.33 (2.58)	15.34 (6.9)
09Q2	1.84	14.62	72.48	11.05	12.98	10.67 (0)	11.35 (0)	12.24 (0)	13.49 (1.04)	15.53 (5.34)
09Q3	2.98	15.83	68.45	12.73	13.26	10.07 (0)	10.71 (0)	11.55 (0)	13.56 (0.6)	14.67 (2.87)
09Q4	4.02	18.4	66.32	11.26	12.94	9.94 (0)	10.41 (0)	11 (0)	11.76 (0)	12.91 (0)
10Q1	3.67	17.66	67.07	11.6	13.17	9.92 (0)	10.39 (0)	10.96 (0)	11.71 (0)	12.82 (0)
10Q2	2.81	17.98	66.72	12.5	13.49	9.93 (0)	10.4 (0)	10.98 (0)	11.73 (0)	12.85 (0)
10Q3	2.39	19.44	66.46	11.7	13.41	10.06 (0)	10.55 (0)	11.15 (0)	11.95 (0)	13.15 (0)
10Q4	3.11	17.35	68.17	11.37	13.38	9.82 (0)	10.31 (0)	10.93 (0)	11.73 (0)	12.93 (0)
11Q1	3.64	17.43	67.76	11.17	11.25	9.78 (0)	10.28 (0)	10.89 (0)	11.69 (0.85)	12.89 (3.22)
11Q2	3.94	17.89	67.27	10.9	11.41	9.77 (0)	10.26 (0)	10.88 (0)	11.69 (0.53)	12.89 (2.92)
11Q3	3.19	17.58	68.85	10.38	11.7	9.89 (0)	10.39 (0)	11.01 (0)	11.82 (0.23)	13.03 (2.64)
11Q4	2.56	17.27	70.35	9.83	11.34	10.83 (0)	11.42 (0.16)	12.16 (1.65)	13.17 (3.72)	14.75 (7.07)

Table 12 The parameters of bond index (left-hand side) and market index (right-hand side)

Years	α	β	c	δ	Years	α	β	c	δ
05	2	0	0.02	0.002	05	2	0	0.033	0.003
06	2	0	0.02	0.002	06	2	0	0.033	0.003
07	2	0	0.02	0.002	07	2	0	0.032	0.003
08	2	0	0.017	0.002	08	2	0	0.033	0.002
09	2	0	0.024	0.002	09	2	0	0.034	0.002
10	2	0	0.024	0.002	10	2	0	0.035	0.003
11	2	0	0.025	0.003	11	2	0	0.034	0.002

Table 13 The estimations of required capital ratios of the Bank of America under normal distribution, all units are in percentage except raising capital is in billion. w_0 is the weight of risk-free assets, the column α is capital ratios in the balance sheet of the bank. Columns 3%, 2.5%, 2%, 1.5%, and 1% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	α	3%	2.5%	2%	1.5%	1%
06Q1	12.85	24.02	45.82	17.31	9.41	9.23 (0)	9.54 (2.03)	9.9 (7.55)	10.35 (14.42)	10.94 (23.71)
06Q2	12.49	23.4	46.81	17.3	8.83	9.24 (6.49)	9.54 (11.41)	9.91 (17.25)	10.35 (24.54)	10.95 (34.38)
06Q3	12.06	20.92	47.28	19.73	9.2	9.02 (0)	9.32 (1.88)	9.67 (7.53)	10.11 (14.56)	10.68 (24.06)
06Q4	12.37	21.14	48.81	17.68	9.24	9.26 (0.24)	9.56 (5.13)	9.92 (10.94)	10.36 (18.18)	10.94 (27.96)
07Q1	11.84	20.91	49.62	17.63	8.93	9.31 (6.22)	9.61 (11.31)	9.97 (17.37)	10.41 (24.92)	11 (35.12)
07Q2	11.43	20.92	50.7	16.95	8.84	9.4 (9.5)	9.71 (14.76)	10.07 (21.02)	10.52 (28.82)	11.12 (39.36)
07Q3	11.08	20.34	51.35	17.23	8.75	9.41 (11.42)	9.72 (16.86)	10.08 (23.32)	10.53 (31.37)	11.13 (42.24)
07Q4	10.45	21.71	52.13	15.71	8.53	9.49 (18.22)	9.8 (24.25)	10.17 (31.41)	10.63 (40.33)	11.23 (52.39)
08Q1	9.61	24.78	51.23	14.38	8.97	9.71 (14.3)	10.03 (20.59)	10.41 (28.06)	10.88 (37.38)	11.5 (49.98)
08Q2	8.79	26.95	50.94	13.31	9.44	9.91 (8.96)	10.24 (15.34)	10.63 (22.92)	11.11 (32.37)	11.75 (45.15)
08Q3	7.35	26.29	51.73	14.63	8.77	9.91 (23.22)	10.24 (30.07)	10.63 (38.21)	11.11 (48.36)	11.75 (62.08)
08Q4	6.76	25.6	51.59	16.05	9.72	9.14 (0)	9.44 (0)	9.8 (1.61)	10.24 (10.54)	10.82 (22.6)
09Q1	14.82	23.48	43.24	18.46	10.33	8.4 (0)	8.67 (0)	8.98 (0)	9.38 (0)	9.9 (0)
09Q2	15.92	22.4	43.22	18.47	11.33	8.25 (0)	8.52 (0)	8.83 (0)	9.21 (0)	9.73 (0)
09Q3	16.74	22.85	41.49	18.93	11.46	8.22 (0)	8.48 (0)	8.79 (0)	9.17 (0)	9.68 (0)
09Q4	15.53	23.93	41.65	18.9	10.41	9.85 (0)	10.17 (0)	10.55 (3.52)	11.02 (15.26)	11.64 (31.13)
10Q1	16.34	23.83	41.9	17.93	9.87	9.86 (0)	10.18 (8.23)	10.56 (18.2)	11.03 (30.63)	11.65 (47.42)
10Q2	18.46	23.63	40.84	17.07	9.91	9.72 (0)	10.04 (3.31)	10.41 (13.17)	10.87 (25.45)	11.48 (42.05)
10Q3	18.76	25.4	40.23	15.61	9.92	9.87 (0)	10.19 (7.12)	10.57 (17.07)	11.04 (29.46)	11.66 (46.21)
10Q4	16.01	25.51	42.12	16.36	10.46	10.08 (0)	10.41 (0)	10.81 (9)	11.3 (21.61)	11.96 (38.66)
11Q1	16.46	25.12	41.39	17.02	10.19	9.96 (0)	10.29 (2.58)	10.69 (12.6)	11.17 (25.08)	11.82 (41.96)
11Q2	17.58	24.63	41.85	15.94	9.86	9.94 (2.04)	10.28 (10.41)	10.67 (20.38)	11.15 (32.8)	11.79 (49.59)
11Q3	16.26	25.62	42.65	15.47	10.39	10.14 (0)	10.48 (2.08)	10.88 (12.08)	11.37 (24.55)	12.03 (41.4)
11Q4	16.93	24.6	43.85	14.62	10.8	10.08 (0)	10.42 (0)	10.83 (0.71)	11.32 (12.7)	11.99 (28.89)

Table 14 The estimations of required capital ratios of the Wells Fargo under normal distribution, all units are in percentage except raising capital is in billion. w_0 is the weight of risk-free assets, the column cr is capital ratios in the balance sheet of the bank. Columns 4.5%, 4%, 3.5%, 3%, and 2.5% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	cr	4.5%	4%	3.5%	3%	2.5%
06Q1	3.69	11.49	70.46	14.36	8.52	9.65 (6.14)	9.9 (7.54)	10.18 (9.08)	10.49 (10.83)	10.85 (12.85)
06Q2	3.89	15.26	67.48	13.36	8.39	9.55 (6.4)	9.8 (7.8)	10.07 (9.35)	10.38 (11.1)	10.73 (13.12)
06Q3	3.45	11.4	71.2	13.94	9.28	9.71 (2.31)	9.97 (3.68)	10.25 (5.2)	10.56 (6.92)	10.92 (8.9)
06Q4	3.96	9.41	72.86	13.77	9.52	10.08 (3.01)	10.34 (4.4)	10.62 (5.95)	10.94 (7.69)	11.3 (9.71)
07Q1	3.52	10.07	73.03	13.38	9.49	10.11 (3.33)	10.37 (4.74)	10.65 (6.31)	10.97 (8.08)	11.34 (10.12)
07Q2	3.3	14.03	69.36	13.31	8.76	9.89 (6.78)	10.15 (8.32)	10.43 (10.03)	10.74 (11.96)	11.1 (14.19)
07Q3	3.04	11.27	71.04	14.65	8.7	9.96 (7.66)	10.21 (9.23)	10.49 (10.98)	10.81 (12.96)	11.17 (15.24)
07Q4	3.04	13.53	70.32	13.12	8.28	9.86 (10.11)	10.11 (11.76)	10.39 (13.59)	10.71 (15.66)	11.07 (18.05)
08Q1	2.91	14.97	69.06	13.06	8.09	9.8 (11.3)	10.05 (13)	10.33 (14.88)	10.64 (17.01)	11 (19.46)
08Q2	2.9	16.26	68.59	12.24	7.87	9.82 (13.13)	10.07 (14.87)	10.35 (16.81)	10.66 (19)	11.02 (21.53)
08Q3	3.37	15.12	67.9	13.62	7.54	9.7 (14.88)	9.95 (16.64)	10.23 (18.59)	10.53 (20.8)	10.89 (23.34)
08Q4	5.59	15.5	66.44	12.48	7.57	8.41 (12.12)	8.63 (15.21)	8.86 (18.64)	9.13 (22.51)	9.43 (26.98)
09Q1	3.17	17.29	67.38	12.16	8.33	8.6 (3.91)	8.82 (7.03)	9.07 (10.48)	9.34 (14.38)	9.65 (18.87)
09Q2	2.85	19.02	65.88	12.26	9.45	8.63 (0)	8.85 (0)	9.1 (0)	9.37 (0)	9.69 (3.32)
09Q3	2.83	18.24	66.52	12.42	10.49	8.62 (0)	8.84 (0)	9.08 (0)	9.35 (0)	9.67 (0)
09Q4	5.46	17.08	64.57	12.88	9.2	10.89 (23.6)	11.17 (27.6)	11.48 (32.04)	11.82 (37.05)	12.22 (42.84)
10Q1	5.76	16.81	65.07	12.36	9.66	10.93 (17.44)	11.21 (21.37)	11.52 (25.74)	11.87 (30.67)	12.26 (36.37)
10Q2	7.18	16.47	64.25	12.09	9.9	10.82 (12.56)	11.09 (16.43)	11.4 (20.73)	11.75 (25.59)	12.14 (31.19)
10Q3	5.72	18.25	63.87	12.17	10.25	10.87 (8.51)	11.15 (12.38)	11.46 (16.68)	11.81 (21.55)	12.21 (27.16)
10Q4	7.61	17.43	62.66	12.31	10.17	10.75 (8.27)	11.03 (12.29)	11.35 (16.75)	11.69 (21.78)	12.09 (27.6)
11Q1	8.84	17.73	61.36	12.07	10.84	10.62 (0)	10.9 (0.83)	11.21 (5.13)	11.55 (9.99)	11.95 (15.6)
11Q2	8.93	18.71	60.63	11.73	10.95	10.6 (0)	10.88 (0)	11.18 (3.32)	11.53 (8.22)	11.92 (13.88)
11Q3	8.29	19.95	60.04	11.72	10.67	10.61 (0)	10.89 (3.2)	11.2 (7.73)	11.54 (12.83)	11.93 (18.73)
11Q4	4.86	22.53	60.89	11.73	10.78	10.75 (0)	11.04 (3.83)	11.36 (8.6)	11.72 (13.98)	12.13 (20.19)

Table 15 The estimations of required capital ratios of the Citigroup under normal distribution, all units are in percentage except raising capital is in billion. w_0 is the weight of risk-free assets, the column α is capital ratios in the balance sheet of the bank. Columns 4.5%, 4%, 3.5%, 3%, and 2.5% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	α	4.5%	4%	3.5%	3%	2.5%
06Q1	18.55	22.69	40.02	18.75	7.21	7.98 (13.27)	8.18 (16.71)	8.4 (20.52)	8.64 (24.82)	8.92 (29.78)
06Q2	18.11	21.7	40.95	19.24	7.1	7.94 (14.95)	8.14 (18.45)	8.36 (22.34)	8.6 (26.72)	8.88 (31.78)
06Q3	18.27	23.83	39.15	18.75	6.75	8.06 (24.84)	8.26 (28.7)	8.48 (32.97)	8.73 (37.8)	9.01 (43.36)
06Q4	18.67	24.64	37.59	19.09	6.36	8.08 (35.34)	8.28 (39.49)	8.5 (44.09)	8.74 (49.28)	9.03 (55.27)
07Q1	18.47	24.55	36.13	20.84	6.04	7.99 (42.86)	8.19 (47.25)	8.4 (52.13)	8.65 (57.62)	8.92 (63.96)
07Q2	20.25	21.37	35.44	22.95	5.75	7.61 (44.76)	7.8 (49.25)	8 (54.23)	8.23 (59.86)	8.49 (66.33)
07Q3	20.36	21.9	35.7	22.04	5.39	7.68 (58.45)	7.86 (63.3)	8.07 (68.68)	8.3 (74.75)	8.56 (81.74)
07Q4	17.45	20.19	38.5	23.86	5.19	7.64 (57.99)	7.83 (62.48)	8.03 (67.47)	8.26 (73.08)	8.52 (79.56)
08Q1	15.6	27.3	40.4	16.7	5.83	8.45 (63)	8.66 (68.23)	8.89 (74.04)	9.16 (80.58)	9.46 (88.13)
08Q2	15.85	27.14	40	17.01	6.49	8.41 (43.9)	8.62 (48.82)	8.85 (54.28)	9.11 (60.44)	9.41 (67.54)
08Q3	17.92	25.7	38.89	17.5	6.15	8.18 (45.25)	8.38 (49.87)	8.6 (55)	8.85 (60.78)	9.14 (67.44)
08Q4	19.79	27.37	36.91	15.93	7.31	7.96 (13.73)	8.15 (17.86)	8.37 (22.44)	8.61 (27.6)	8.88 (33.55)
09Q1	20.31	26.88	36.85	15.96	8.01	7.89 (0)	8.08 (1.46)	8.29 (5.68)	8.53 (10.43)	8.8 (15.91)
09Q2	21.27	27.45	35.42	15.85	8.32	7.91 (0)	8.1 (0)	8.32 (0)	8.55 (4.63)	8.83 (10.19)
09Q3	23.71	27.2	33.54	15.55	7.55	7.8 (5.15)	7.99 (9.03)	8.2 (13.34)	8.43 (18.2)	8.7 (23.8)
09Q4	22.32	29.88	32.28	15.51	8.32	9.04 (14.68)	9.26 (19.31)	9.51 (24.45)	9.79 (30.25)	10.1 (36.93)
10Q1	21.19	29.16	35.89	13.76	7.68	9.2 (33.52)	9.43 (38.66)	9.68 (44.37)	9.97 (50.81)	10.29 (58.24)
10Q2	21.48	28.84	35.95	13.72	8.12	9.16 (22.27)	9.39 (27.19)	9.64 (32.66)	9.92 (38.82)	10.25 (45.94)
10Q3	20.93	31.24	34.74	13.09	8.33	9.38 (23.11)	9.62 (28.32)	9.88 (34.1)	10.17 (40.62)	10.5 (48.15)
10Q4	22.82	29.82	34.17	13.19	8.66	9.12 (9.56)	9.35 (14.49)	9.61 (19.97)	9.89 (26.15)	10.22 (33.28)
11Q1	23.22	29.85	33.14	13.79	8.9	9.05 (3.1)	9.28 (8.05)	9.53 (13.55)	9.82 (19.75)	10.14 (26.91)
11Q2	23.91	28.76	33.64	13.68	9.13	8.94 (0)	9.17 (0.82)	9.42 (6.24)	9.7 (12.34)	10.02 (19.38)
11Q3	24.74	28.85	33.5	12.9	9.26	8.94 (0)	9.17 (0)	9.42 (3.28)	9.7 (9.31)	10.02 (16.26)
11Q4	24.57	27.58	35.27	12.57	9.58	8.89 (0)	9.12 (0)	9.37 (0)	9.65 (1.37)	9.97 (8.05)

Table 16 The estimations of required capital ratios of the Regions Financial under normal distribution, all units are in percentage except raising capital is in billion. w_0 is the weight of risk-free assets, the column α is capital ratios in the balance sheet of the bank. Columns 3%, 2.5%, 2%, 1.5%, and 1% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	α	3%	2.5%	2%	1.5%	1%
06Q1	3.51	14.43	70.33	11.73	12.6	10.61 (0)	10.97 (0)	11.4 (0)	12.25 (0)	12.63 (0.03)
06Q2	3.57	14.33	70.78	11.31	12.43	10.65 (0)	11.01 (0)	11.44 (0)	12.29 (0)	12.68 (0.24)
06Q3	3.46	15.44	69.98	11.12	12.7	10.63 (0)	10.99 (0)	11.42 (0)	12.27 (0)	12.65 (0)
06Q4	3.22	13.4	68.86	14.52	14.44	10.66 (0)	11.02 (0)	11.44 (0)	12.27 (0)	12.65 (0)
07Q1	3.02	13.86	68.45	14.67	14.71	10.64 (0)	11 (0)	11.42 (0)	12.25 (0)	12.62 (0)
07Q2	2.93	13.24	68.89	14.94	14.31	10.66 (0)	11.02 (0)	11.44 (0)	12.27 (0)	12.64 (0)
07Q3	2.63	12.75	68.19	16.42	14.36	10.57 (0)	10.93 (0)	11.34 (0)	12.17 (0)	12.54 (0)
07Q4	3.49	12.71	67.29	16.51	14.05	10.37 (0)	10.72 (0)	11.13 (0)	11.94 (0)	12.3 (0)
08Q1	2.89	13.74	66.51	16.86	13.88	10.34 (0)	10.69 (0)	11.09 (0)	11.9 (0)	12.26 (0)
08Q2	2.89	12.75	67.65	16.7	13.64	10.41 (0)	10.76 (0)	11.17 (0)	11.98 (0)	12.35 (0)
08Q3	2.57	12.4	68.23	16.8	13.66	10.46 (0)	10.81 (0)	11.22 (0)	12.04 (0)	12.41 (0)
08Q4	7.51	13.34	66.35	12.81	11.5	8.97 (0)	9.27 (0)	9.62 (0)	10.31 (0)	10.63 (0)
09Q1	3.62	15.54	67.55	13.29	12.04	9.22 (0)	9.53 (0)	9.89 (0)	10.62 (0)	10.94 (0)
09Q2	5.9	14.24	67.11	12.74	13.32	9.09 (0)	9.4 (0)	9.75 (0)	10.46 (0)	10.78 (0)
09Q3	5.97	15.61	65.43	13	13.39	9.06 (0)	9.36 (0)	9.71 (0)	10.42 (0)	10.74 (0)
09Q4	5.64	18.55	62.61	13.2	12.76	11.68 (0)	12.07 (0)	12.53 (0)	13.45 (1.13)	13.86 (1.82)
10Q1	5	18.09	62.7	14.2	13.04	11.66 (0)	12.06 (0)	12.52 (0)	13.43 (0.62)	13.84 (1.28)
10Q2	5.47	18.4	62.02	14.11	13.09	11.6 (0)	11.99 (0)	12.45 (0)	13.36 (0.42)	13.77 (1.06)
10Q3	5.15	18.64	62.06	14.15	13.05	11.62 (0)	12.01 (0)	12.47 (0)	13.38 (0.52)	13.79 (1.16)
10Q4	5.22	18.06	61.35	15.36	12.83	11.58 (0)	11.98 (0)	12.44 (0)	13.37 (0.82)	13.78 (1.46)
11Q1	5.54	19.31	60.55	14.61	12.81	11.56 (0)	11.96 (0)	12.42 (0)	13.35 (0.83)	13.76 (1.47)
11Q2	6.07	18.8	60.55	14.58	13.1	11.53 (0)	11.92 (0)	12.39 (0)	13.31 (0.32)	13.72 (0.95)
11Q3	6.36	19.79	59.78	14.07	13.5	11.5 (0)	11.9 (0)	12.36 (0)	13.28 (0)	13.69 (0.28)
11Q4	5.69	20.05	59.93	14.33	13.19	11.42 (0)	11.82 (0)	12.29 (0)	13.22 (0.06)	13.64 (0.67)

Table 17 The estimations of required capital ratios of the SunTrust under normal distribution, all units are in percentage except raising capital is in billion. w_0 is the weight of risk-free assets, the column cr is capital ratios in the balance sheet of the bank. Columns 3%, 2.5%, 2%, 1.5%, and 1% are estimated required capital ratios (raising capital) with corresponding bankruptcy risk. The value in the shaded cell is the closest estimation

Years	w_0	S/A	L/A	O/A	cr	3%	2.5%	2%	1.5%	1%
06Q1	2.99	16.2	70.69	10.12	9.59	10.74 (2.31)	11.11 (3.06)	11.55 (3.95)	12.08 (5.06)	12.79 (6.57)
06Q2	2.86	15.65	71.77	9.72	9.62	10.82 (2.43)	11.19 (3.2)	11.62 (4.11)	12.16 (5.25)	12.88 (6.79)
06Q3	2.87	15.49	71.9	9.74	10.15	10.82 (1.37)	11.19 (2.14)	11.63 (3.06)	12.17 (4.2)	12.89 (5.75)
06Q4	2.91	14.74	72.57	9.78	9.78	11.1 (2.71)	11.47 (3.49)	11.91 (4.41)	12.46 (5.57)	13.18 (7.13)
07Q1	2.56	17.42	69.72	10.3	9.64	10.95 (2.74)	11.32 (3.53)	11.75 (4.46)	12.29 (5.63)	13 (7.2)
07Q2	3.01	14.07	72.21	10.71	9.63	11.03 (2.84)	11.4 (3.6)	11.84 (4.51)	12.38 (5.65)	13.1 (7.19)
07Q3	2.93	12.73	72.97	11.36	10.18	11.06 (1.73)	11.43 (2.47)	11.86 (3.36)	12.4 (4.46)	13.12 (5.95)
07Q4	3.14	14.09	72.33	10.44	10.05	10.93 (1.77)	11.3 (2.53)	11.73 (3.42)	12.27 (4.54)	12.98 (6.05)
08Q1	2.94	13.57	72.61	10.88	10.3	10.94 (1.29)	11.31 (2.04)	11.74 (2.93)	12.28 (4.04)	12.99 (5.54)
08Q2	3.11	12.79	73.61	10.49	10.1	11.01 (1.81)	11.38 (2.56)	11.82 (3.45)	12.36 (4.56)	13.08 (6.06)
08Q3	2.62	11.84	74.88	10.66	10.27	11.11 (1.65)	11.49 (2.39)	11.93 (3.28)	12.47 (4.39)	13.19 (5.88)
08Q4	3.51	14.43	68.49	13.58	11.84	9.22 (0)	9.53 (0)	9.89 (0)	10.34 (0)	10.94 (0)
09Q1	3.94	13.93	71.76	10.37	12.08	9.44 (0)	9.76 (0)	10.13 (0)	10.59 (0)	11.21 (0)
09Q2	1.84	14.62	72.48	11.05	12.98	9.55 (0)	9.87 (0)	10.25 (0)	10.71 (0)	11.34 (0)
09Q3	2.98	15.83	68.45	12.73	13.26	9.31 (0)	9.62 (0)	9.99 (0)	10.45 (0)	11.05 (0)
09Q4	4.02	18.4	66.32	11.26	12.94	12.11 (0)	12.52 (0)	13 (0.13)	13.6 (1.33)	14.39 (2.95)
10Q1	3.67	17.66	67.07	11.6	13.17	12.16 (0)	12.57 (0)	13.05 (0)	13.65 (0.95)	14.44 (2.56)
10Q2	2.81	17.98	66.72	12.5	13.49	12.14 (0)	12.55 (0)	13.03 (0)	13.62 (0.26)	14.42 (1.84)
10Q3	2.39	19.44	66.46	11.7	13.41	12.2 (0)	12.61 (0)	13.1 (0)	13.69 (0.57)	14.49 (2.2)
10Q4	3.11	17.35	68.17	11.37	13.38	12.37 (0)	12.8 (0)	13.3 (0)	13.92 (1.08)	14.74 (2.76)
11Q1	3.64	17.43	67.76	11.17	11.25	12.32 (2.09)	12.75 (2.93)	13.25 (3.93)	13.86 (5.18)	14.68 (6.87)
11Q2	3.94	17.89	67.27	10.9	11.41	12.29 (1.72)	12.71 (2.56)	13.21 (3.56)	13.82 (4.81)	14.64 (6.51)
11Q3	3.19	17.58	68.85	10.38	11.7	12.47 (1.5)	12.9 (2.36)	13.4 (3.38)	14.03 (4.66)	14.86 (6.39)
11Q4	2.56	17.27	70.35	9.83	11.34	12.48 (2.29)	12.92 (3.2)	13.44 (4.28)	14.08 (5.64)	14.94 (7.47)

banks needed to raise their capital before 2008, it is the signal that the banking system will be in trouble in the crisis period. All banks decreased lending ratios in their portfolio significantly after the bank regulator's announcement, even though the purpose of government bailout is to prop up the capital and new loans. The "externality" issue of the banking system is a major reason such that the government-regulated banks should reach an adequate capital to prevent banks run. In this chapter, we show that the required capital ratios reduced significantly after increasing the weight of risk-free assets. This paper gives rational reasons that banks take these actions. In this chapter, total assets are simply classified into risk-free asset, loans, and securities. They are log-stable Paretian distributed with different characteristic exponents. However, this assumption neglects the correlation of assets. The diversification effect will be underestimated. Therefore, banks may hold excessive capital which is not necessary, especially for the normal period. The copula method will be one of the ways to tackle this issue in our future research (Tables 12, 13, 14, 15, 16, and 17).

References

- Acharya, V., R.F. Engle, and M. Richardson. 2012. Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review* 102 (3): 59–64.
- Acharya, V., I. Drechsler, and P. Schnabl. 2014. A pyrrhic victory? Bank bailouts and sovereign credit risk. *The Journal of Finance* 69 (6): 2689–2739.
- Akhigbe, A., J. Madura, and M. Marciniak. 2012. Bank capital and exposure to the financial crisis. *Journal of Economics and Business* 64 (5): 377–392.
- Allen, F., and D. Gale. 2001. Bubbles and crises. *Economic Journal* 110 (460): 236–255.
- Braggion, F., and L. Moore. 2011. Dividend policies in an unregulated market: The London stock exchange, 1895–1905. *The Review of Financial Studies* 24 (9): 2935–2973.
- Brownlees, C. T., and R. F. Engle. 2011. Volatility, correlation and tails for systemic risk measurement. *Working paper*.
- Buser, D., A. Chen, and E. Kane. 1981. Federal deposit insurance, regulatory policy, and optimal bank capital. *Journal of Finance* 36 (1): 51–60.
- Castiglionesi, F., F. Feriozzi, G. Loranth, and L. Pelizzon. 2014. Liquidity coinsurance and bank capital. *Journal of Money, Credit, and Banking* 46 (2–3): 409–443.
- Cornett, M., J. McNutt, P. Strahan, and H. Tehranian. 2011. Liquidity risk management and credit supply in the financial crisis. *Journal of Financial Economics* 101 (2): 297–312.
- Coronel-Brizio, H.F., and A.R. Hernandez-Montoya. 2005. On fitting the pareto levy distribution to stock market index data: Selecting a suitable cutoff value. *Physica A* 354 (15): 437–449.
- Correa, R., K.-H. Lee, H. Sapriza, and G.A. Suarez. 2014. Sovereign credit risk, banks' government support, and bank stock returns around the world. *Journal of Money, Credit and Banking* 46: 93–121.
- DuMouchel, W.H. 1983. Estimating the stable index α in order to measure tail thickness: A critique. *The Annals of Statistics* 11 (4): 1019–1031.
- Fama, E.F., and K.R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.
- Flannery, M.J. 2014. Maintaining adequate bank capital. *Journal of Money, Credit and Banking* 46: 157–180.
- Freixas, X., A. Martin, and D. Skeie. 2011. Bank liquidity, interbank markets, and monetary policy. *Review of Financial Studies* 24 (8): 2656–2692.
- Galai, D., and Z. Wiener. 2012. Credit risk spreads in local and foreign currencies. *Journal of Money, Credit and Banking* 44 (5): 883–901.

- Garel, A., and A. Petit-Romec. 2017. Bank capital in the crisis: It's not just how much you have but who provides it. *Journal of Banking & Finance* 75: 152–166.
- Hovakimian, A., and E.J. Kane. 2000. Effectiveness of capital regulation at U.S. commercial banks, 1985 to 1994. *The Journal of Finance* 55 (1): 451–468.
- Ivashina, V., and D. Scharfstein. 2010. Bank lending during the financial crisis of 2008? *Journal of Financial Econometrics* 97 (3): 319–338.
- Keister, T., and J. McAndrews. 2009. Why are banks holding so many excess reserves? *FRB New York Current Issues in Economics and Finance* 15 (8).
- Liu, J., J. Pan, and T. Wang. 2005. An equilibrium model of rare-event premia and its implication for option smirks. *The Review of Financial Studies* 18 (1): 131–164.
- McCulloch, J.H. 1997. Measuring tail thickness to estimate the stable index α : A critique. *Journal of Business & Economic Statistics* 15 (1): 74–81.
- McCulloch, J. H., and Lee, S. H. 2008. *Estimation of risk neutral measures using the generalized two-factor log-stable option pricing model*. Working paper, Ohio State University.
- Merton, R.C. 1969. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics* 51 (3): 247–257.
- Miller, M.H., and K. Rock. 1985. Dividend policy under asymmetric information. *The Journal of Finance* 40 (4): 1031–1051.
- Nolan, J. P. 2001. *Maximum likelihood estimation and diagnostics for stable distributions*. Technical report.
- Nolan, J.P., and D. Ojeda-Revah. 2013. Linear and nonlinear regression with stable errors. *Journal of Econometrics* 172 (2): 186–194.
- Samuelson, P.A. 1967. Efficient portfolio selection for Pareto-Lévy investments. *The Journal of Financial and Quantitative Analysis* 2 (2): 107–122.
- U, S.C., J. So, D. Ding, and L. Liu. 2016. An efficient Fourier expansion method for the calculation of value-at-risk: Contributions of extra-ordinary risks. *International Journal of Financial Engineering* 03: 1650006.
- So, J. C., and S. C. U. 2012. Value at risk and stable Paretian distributions. *Conference paper, Decision Science Institute 2012 Annual Meeting*, San Francisco.
- Yang, J., and Y. Zhou. 2013. Credit risk spillovers among financial institutions around the global credit crisis: Firm-level evidence. *Management Science* 59 (10): 2343–2359.



The Economics of and Accounting for Lease Transactions

86

Cheng-Few Lee and Ryan McDonough

Abstract

This chapter explores a company's decision to enter into a lease contract to finance capital investments. We discuss the motivations behind the lease-versus-buy tradeoff in terms of both the economics of and accounting for lease financing arrangements. Throughout our discussion, we reference academic research related to the lease-versus-buy decision, the factors influencing the extent of leasing activity, the economic consequences of lease financing, and the role of lease-related accounting standards. We also provide a detailed review of the provisions in the current lease accounting standards that were recently issued by the Financial Accounting Standards Board and the International Accounting Standards Board.

Keywords

ASC 842 · IFRS 16 · Investment · Finance leases · Lease financing · Lease-versus-buy · Operating leases

JEL Classification

E22 · G32 · G38 · M41 · M48

C.-F. Lee

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: cflee@business.rutgers.edu

R. McDonough (✉)

Department of Accounting and Information Systems, Rutgers University, Piscataway, NJ, USA

e-mail: mcdonough@business.rutgers.edu

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,

https://doi.org/10.1007/978-3-030-91231-4_88

2019

86.1 Introduction

Capital expenditure decisions are among the most fundamental and important decisions made by corporate managers. When making investment decisions, corporate managers are confronted with the choice to either purchase an asset or, alternatively, enter into a lease contract to finance the use of an asset for some or all of the asset's economic life. This chapter examines companies' decisions to enter into lease transactions, with an emphasis on the economic theory underpinning lease financing as well as the accounting and financial reporting implications that have evolved in recent years.

Leases are ubiquitous in practice. In addition to the economic motivations for lease financing, accounting and reporting considerations have historically influenced the benefits and costs associated with leases. Indeed, a flourishing stream of academic research has sought to explain the existence of leases and has explored the economics of leasing and the impact of accounting requirements on lease structuring (e.g., Johnson and Lewellen 1972; Gordon 1974; Miller and Upton 1976; Ang and Peterson 1984; Imhoff and Thomas 1988; Eisfeldt and Rampini 2009; Caskey and Ozel 2019). The finance literature has generally focused on explaining the prevalence of lease transactions through the lens of borrowing capacity (e.g., Sharpe and Nguyen 1995; Eisfeldt and Rampini 2009), tax considerations (e.g., Miller and Upton 1976; Myers et al. 1976; Graham et al. 1998), and operational flexibility (e.g., Gavazza 2011). Meanwhile, accounting scholars have generally focused on the impact of accounting and financial reporting requirements in terms of both companies' incentives to structure leases to be off-balance-sheet and the capital market impacts of such structuring decisions (see, e.g., Spencer and Webb 2015). We discuss this body of research throughout this chapter.

In the next section, we define the nature of a lease and discuss the use of lease arrangements across industries. We then provide an overview of the economic theory and assumptions that help explain the existence of leasing as a widely used investment financing option. The fourth section focuses on the accounting standards issued by the Financial Accounting Standards Board and the International Accounting Standards Board that have recently ushered in a new era of accounting and reporting for leases. We discuss how the recent changes in accounting and financial reporting requirements have potentially shifted the expected costs and benefits of some lease arrangements. Although we focus primarily on generally accepted accounting principles (GAAP) in the United States, we also provide a brief discussion of international financial reporting standards (IFRS). The last section concludes this chapter.

86.1.1 Nature and Use

86.1.1.1 General Nature of a Lease Arrangement

Companies lease assets such as land, agricultural products, buildings, plants, retail and office space, and equipment. A lease transaction is guided by a contractual

agreement that specifies the rights and obligations of the contracting parties – the lessor and the lessee. The lessor commonly owns the asset to be leased. The lessee has an obligation to make payments to the lessor in exchange for the exclusive right to control the use of the lessor's physical asset (e.g., property, plant, or equipment) for a specified period of time (the lease term). The lease term is generally less than the expected economic life of the underlying leased asset; otherwise, the transfer of possession of the asset would generally be considered a sale. At the end of the lease term, the underlying leased asset is typically returned to the lessor. The lessor could then, for example, redeploy the asset to another lessee or to a buyer.

86.1.1.2 Use of Leases

Leases are used across the world and, historically, have been used for at least as long as contractual arrangements have been recorded (e.g., Ellickson and Thorland 1995, p. 369; Spencer and Webb 2015, p. 1000). Although the use of leases spans industries, companies in industries such as airlines, retail, and telecommunications tend to rely relatively more on leases as a financing option than companies in other industries do. Present day lease transactions include, for example, retail space leased by companies such as Walgreens Boots Alliance Inc. and CVS Health Corporation. AT&T Inc. and Verizon Communications Inc. both have lease arrangements that cover assets such as towers, retail stores, and buildings. Amazon Inc., FedEx Corporation, and United Parcel Service Inc. lease, for example, aircraft, distribution facilities, and vehicles. Companies in the airline industry, such as United Continental Holdings and Delta Airlines Inc., lease aircraft, buildings, and other assets. According to prior research, slightly more than half of all commercial aircraft are leased (Gavazza 2010, pp. 68–69).

Securities and Exchange Commission (SEC) Form 10-K filings for companies' 2019 fiscal year-end reveal that more than 70% of publicly traded US companies report a lease liability. Similarly, prior research suggests that 80% of all US companies lease assets (Gavazza 2010, p. 62). For the 2019 fiscal year-end, publicly traded US companies report an average lease liability of nearly \$190 million. However, the average lease liability reported by the 100 publicly traded US companies with the largest lease liabilities is approximately \$5.2 billion. Illustration 88.1 below further highlights the magnitude of lease liabilities for the 25 US companies with the largest lease liabilities. As the chart demonstrates, the largest lease liability amounts tend to be reported by companies in the airlines, retail, and telecommunications industries. Note that before 2019, the lease liability reported by the vast majority of companies would have been substantially smaller because companies could previously structure their leases to keep many leases (specifically, operating leases) off their balance sheets. The accounting rules currently in place, which are discussed in Sect. 86.3, require companies to report the vast majority of their leases on their balance sheets, regardless of lease classification.

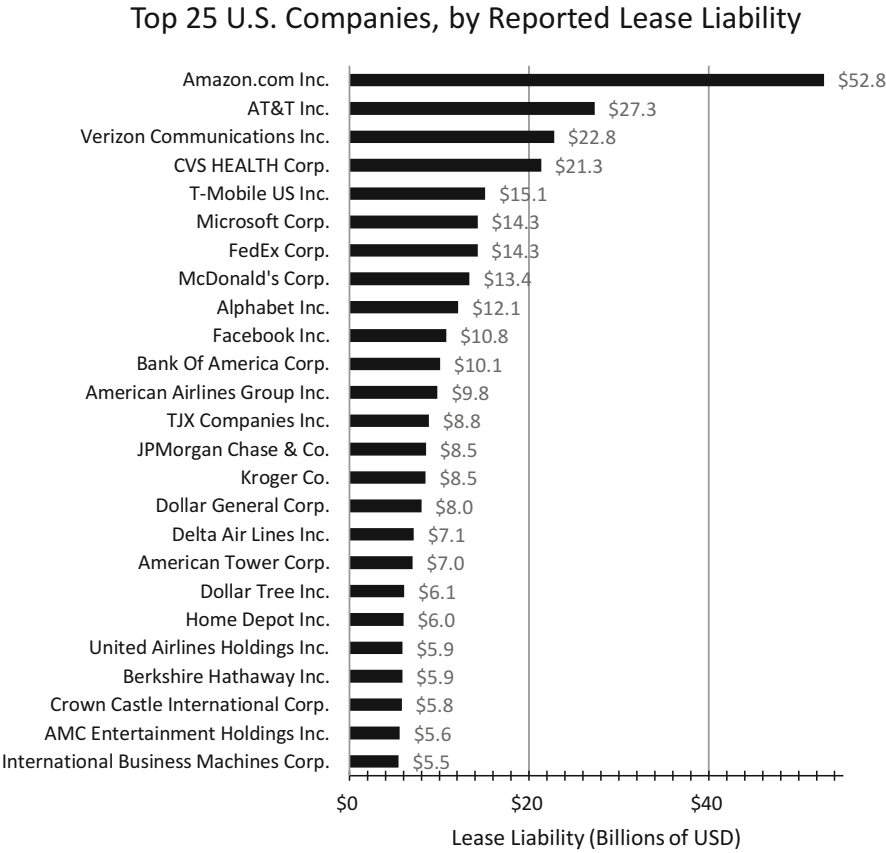


Illustration 88.1 Total lease liability – that is, finance and operating lease liabilities – as of the 2019 fiscal year-end. The amounts are reported in billions of US dollars. (Source: SEC 10-K filings)

86.2 Motivations for Leasing: Theory and Assumptions

There are several factors that can make leasing an attractive and pervasive financing option for many companies. This section discusses the primary motivations for entering into a lease arrangement, including considerations related to transaction costs, operational flexibility, financing capacity, tax consequences, and financial reporting outcomes. It should be noted, however, that the discussion that follows is not meant to be a blanket endorsement of leasing. There are, of course, many situations in which it makes more sense for a company to purchase – whether using internal funds, debt, or equity – than to lease an asset. The discussion in this section is only meant to highlight considerations that can make leasing comparatively more attractive than purchasing in certain circumstances, thus explaining the observed widespread use of lease financing.

86.2.1 Financing: Lease Vs. Purchase

A natural question that arises is whether and to what extent there is a difference in the economic substance of purchasing an asset through the issuance of long-term debt and purchasing the right to use an asset by entering into a lease agreement. Generally speaking, the economic substance of these two transactions is identical, although the contractual arrangements and contracting parties will differ to some extent. Indeed, following from the Modigliani-Miller theorem (Modigliani and Miller 1958, 1963), Miller and Upton (1976) assert that under certain fairly strict assumptions, the decision to lease an asset rather than borrowing funds to purchase it should be irrelevant to the company (see also, e.g., Smith and Wakeman 1985). However, the assumptions underlying this irrelevance conclusion likely do not reflect reality. For instance, Miller and Upton's assumptions relating to the existence of perfect capital markets, as well as the absence of asymmetries in transaction costs, default risk, and tax consequences between the leasing and borrowing-to-buy decisions, do not generally comport well with realistic institutional settings. The discussion that follows focuses on nonfinancial-reporting incentives first and then turns to financial reporting incentives that could play a role in leasing decisions and in how leases are structured.

86.2.2 Nonreporting Incentives: Transaction Costs, Flexibility, Borrowing Capacity, and Taxes

A benefit of leasing, particularly in the absence of access to perfect capital markets, is that companies obtain the exclusive right to control the use of an asset at a relatively low cost, given that the lease is for less than the expected useful life of an asset. In contrast, a purchase is for the asset's entire useful life. Although a company could purchase an asset and then subsequently sell it to make the cost of leasing and purchasing equivalent (for a given period of time), this argument relies on the assumption of perfect capital markets and no asymmetry in transaction costs. Since a lessor can specialize in leasing and redeploying assets while lessees specialize in utilizing assets, the overall costs associated with using the asset are reduced – more generally, the lessee and lessor can efficiently separate responsibilities. A lessor, for example, can use its expertise in acquiring land, building management, maintenance, and asset disposal and redeployment. Lessees can focus on their competitive advantages, whether that be in retail, manufacturing, or services industries.

Similar to the benefits associated with specialization, leases allow lessees and lessors to split the residual rights of the leased asset and, thus, the risks associated with owning the asset. For instance, the lessee pays for the right to use the asset for a certain period of time while the lessor retains some of the residual rights. In particular, since the lease is for a period less than the expected useful life of the asset, the lessee returns the asset to the lessor at the end of the lease term, which

represents the remaining residual rights in the asset.¹ This protects a lessee, for instance, against the risk of obsolescence and also affords the lessee more operational flexibility since the lessee will not need to use the asset for its entire useful life, nor will the lessee need to incur the costs associated with disposing of the asset (including identifying a buyer for a partially used asset). Leases may also offer flexibility vis-à-vis the option to buy the asset or renew the lease at the end of the lease term, as well as the option to choose a lease term suitable to the needs of the company or terminate the lease early in a relatively low-cost and straightforward manner. Gavazza (2011) suggests that operational flexibility (e.g., the ability to adjust for changes in capacity needs at relatively low costs) is an important factor explaining the use of lease financing.

Furthermore, leases typically require relatively small (or even no) down payment, which can make it more attractive than a secured loan that may require a relatively large down payment, especially for financially constrained companies. Leases also offer greater protections to lessors – for example, with respect to repossessing an asset when a lessee does not make the required lease payments – than protections available to secured lenders (Sharpe and Nguyen 1995; Eisfeldt and Rampini 2009; Beatty et al. 2010). This bankruptcy protection afforded to lessors effectively expands a lessee's borrowing capacity beyond what would be the case in a world without leases and only secured borrowing.

A potential additional benefit of using a lease to finance the use of an asset is the associated tax advantages (Miller and Upton 1976; Graham et al. 1998). Similar to the bankruptcy protections for lessors, the tax benefits of leasing will vary by jurisdiction. A company that borrows to build or buy an asset may enjoy tax savings from the deduction of interest and depreciation. If Miller and Upton's (1976) assumptions hold, the tax advantage associated with leasing will equal the tax advantage associated with the deduction of interest and depreciation. However, interest and depreciation generally can only be deducted insofar as a company has taxable income. In contrast, lease payments can be used as a tax shield in future periods (i.e., in the case of net operating losses). More generally, leases can be advantageous because they allow the contracting parties to structure the arrangement so that the tax benefits are allocated to taxable companies.

86.2.3 Financial Reporting Incentives

Finally, it should be noted that the US and international accounting rules previously allowed for off-balance-sheet reporting for certain types of leases. In particular, until a few years ago, the characteristics of a lease arrangement and, thus, the classification for accounting purposes, substantially impacted the accounting treatment (i.e.,

¹ Naturally, the lessor will be concerned about the condition of the leased asset at the end of the lease term. As such the lessor may monitor the lessee and include in the lease agreement certain clauses about the minimum residual value of the asset at the end of the lease.

on- vs. off-balance-sheet reporting). As a result, many lease arrangements were structured so that the lease was not reported as a liability and corresponding asset on a lessee's balance sheet, while nearly identical lease arrangements may have resulted in the lease-related asset and liability being recognized on the balance sheet. Structuring leases to be off-balance-sheet may have been considered a benefit to lessees, as evidenced by widespread efforts by many lessee companies to keep leases off their balance sheets (e.g., Imhoff and Thomas 1988; SEC 2005). The potential financial reporting incentives arise from a desire to improve the reported financial condition of a company – for example, in anticipation of a debt or equity issuance, to avoid debt covenant violations, or to maximize compensation or other contracting outcomes. However, some have argued that the financial reporting incentives associated with keeping leases off-balance-sheet (i.e., through window dressing) are likely not of primary importance (e.g., Myers et al. 1976, p. 815; Caskey and Ozel 2019), but rather the decision to structure a lease to keep it off-balance-sheet was likely primarily driven by tax and bankruptcy laws.² Regardless, the new accounting rules, which are discussed next, generally no longer allow for off-balance-sheet reporting of leases.

86.3 Accounting for Lease Transactions Under US GAAP and IFRS

In November 1976, the Financial Accounting Standards Board (FASB) issued Statement of Financial Accounting Standards No. 13, *Accounting for Leases*, which required a lease arrangement to be classified as either an operating lease or a capital lease (FASB 1976).³ An operating lease was not required to be reported on a lessee's balance sheet, which was generally considered a desirable accounting outcome for lessees, as indicated in the prior section (such an arrangement was considered a “true lease” or rental agreement). A lessee would instead record lease (or rent) expense in the income statement and disclose the lease commitment in the notes to the financial statements. In contrast, the FASB concluded that, for capital leases, a lease-related asset and liability should be reported on a lessee's balance sheet because substantially all of the risks and rewards associated with the underlying leased asset were considered to be transferred to the lessee (i.e., such an

²The tax and bankruptcy laws in the U.S. and many other countries maintain relatively similar criteria for classifying a lease as either a “true lease” (basically a rental agreement) or a financed purchase (Eisfeldt and Rampini 2009, p. 1626). This is important because the same general criteria that allowed a company to report a lease off-balance-sheet for financial reporting purposes also resulted in tax benefits and bankruptcy protections that make it difficult to determine whether a company was structuring a lease for a desired financial reporting outcome or for the benefits associated with tax consequences or expanded borrowing capacity.

³FASB Statement No. 13, which is now superseded, was codified as Topic 840, *Leases*. Note that prior to Statement No. 13, leases were only disclosed in the notes to the financial statements.

arrangement was considered a de facto financed acquisition of the underlying leased asset).

The FASB's Statement No. 13 lease accounting requirements (the predecessor standard) were frequently criticized by financial statement users because the rules relied on bright-line tests that were used to determine if a lease should be capitalized or not. These tests, which are, to some extent, still used today, are discussed in the next section. Because of the all-or-nothing approach to lease capitalization, it was relatively easy for companies to structure lease arrangements in a way that a preferred financial reporting outcome could be obtained. This resulted in a proliferation of off-balance-sheet operating lease arrangements (e.g., Cornaggia et al. 2013, p. 349) and a lack of transparency regarding a company's financing activities (e.g., Monson 2001; Rapoport 2013). To correct for this deficiency and to aid in the comparison of financial statements across companies with different financing arrangements (e.g., leasing vs. debt financing), investors and analysts routinely used note disclosures to reconstruct financial statements and financial ratios as if operating leases had been capitalized and properly treated as a liability.⁴ This was commonly referred to as constructive capitalization (e.g., Imhoff et al. 1991, 1997). There has been considerable debate about whether the disclosed operating lease amounts were treated similarly to on-balance-sheet liabilities and appropriately factored into equity prices and credit assessments (e.g., Wilkins and Zimmer 1983; Ely 1995; Dhaliwal et al. 2011; Bratten et al. 2013; Altamuro et al. 2014; Kraft 2015).⁵ However, the consensus generally was that the FASB's lease accounting standard was deficient in that it could result in differing accounting outcomes for otherwise similar transactions.

In response to these shortcomings and following a June 2005 report from the SEC (SEC 2005), the FASB initiated a standard-setting project with the International Accounting Standards Board to develop new lease accounting standards.⁶ The new accounting rules were issued in February 2016 in the FASB's Accounting Standards Update 2016–02, which is codified, along with subsequent amendments, in the Accounting Standards Codification (ASC) as Topic 842, *Leases* (FASB 2016). ASC 842 was effective for publicly traded companies beginning after December 15, 2018. Although the new guidance largely retains the prior classification criteria established in FASB Statement No. 13, the accounting treatment corresponding to the different classifications has converged to generally require on-balance-sheet reporting for the vast majority of leases. These accounting rules are discussed next, beginning with the lease classification requirements.

⁴FASB Statement No. 13 required lessees to disclose certain information about operating lease commitments, including the minimum lease payments due in the current fiscal year, the subsequent five fiscal years, and a lump sum amount reflecting the fiscal years thereafter.

⁵We refer readers to Spencer and Webb (2015) for a review of the literature related to lease accounting from the perspective of lessees.

⁶The SEC, for instance, indicated in its report that it estimated there was outstanding a total of \$1.25 trillion in off-balance-sheet operating lease obligations.

86.3.1 Classification of Leases

The FASB defines a lease arrangement in ASC 842 as a contractual arrangement that confers the right to control the use of an identified asset – that is, property, plant, or equipment – for a specified period of time in exchange for consideration paid to the lessor. However, there are some scope exceptions. ASC 842 does not apply to leases of intangible assets, leases for the exploration or use of natural resources, leases of biological assets, leases of assets under construction, or leases of inventory.

Once the underlying asset is identified and it is determined that a lease exists, a company must classify the lease based on the provisions of the contract.⁷ Under the new accounting requirements set forth in ASC 842, lessees classify leases as operating leases or finance leases at the lease commencement date.⁸ Classification as an operating lease or finance lease by the lessee depends on whether the lease arrangement meets any of the five criteria described in ASC 842-10-25-2. If a lease meets any of the five criteria described in ASC 842-10-25-2 (see below), the lease effectively transfers control of the underlying leased asset to the lessee, and the lease is classified as a finance lease.⁹ If the lease does not meet any of the five criteria, the lessee is considered to have obtained the right to use an asset, but not control of the underlying leased asset. Such leases are classified as operating leases. The five criteria stipulated by ASC 842-10-25-2 that are used by a lessee to determine whether to classify a lease as a finance lease are as follows:

1. **Ownership transfer test.** The lease arrangement confers ownership of the underlying leased asset to the lessee before or at the end of the lease term.
2. **Purchase option test.** The lease arrangement provides the lessee with an option to purchase the underlying leased asset, and it is reasonably certain the lessee will exercise the option.
3. **Lease term test.** The lease term (or duration of the lease) is for the major part of the remaining economic life of the underlying leased asset.

⁷For simplicity, we refer to a lease as if it applies to only a single leased asset. In practice, additional complexities can arise. For example, an arrangement could include several lease components as well as nonlease components. ASC 842 requires that a company allocate the contract consideration to each separate component of the contract. Each lease component would have to be classified and accounted for separately. Nonlease components (e.g., for the provision of services, such as maintenance) would also need to be accounted for separately following the appropriate accounting guidance.

⁸The previous lease accounting rules required that a lessee classify a lease at the inception date (e.g., the date the lease is signed). ASC 842 requires that a lessee determine whether a lease exists at the inception date, while the classification determination is made at the commencement date. The commencement date is the date on which a lessor makes an underlying leased asset available to the lessee.

⁹As noted in the section addressing lessor accounting, all else equal, the lessor would classify the lease as a sales-type lease, reflecting the de facto sale of the asset from the perspective of conferring control of the underlying leased asset to the lessee.

4. **Lease payments test.** The present value of the sum of the lease payments and any residual value guaranteed by the lessee (that is not already part of the lease payments) equals or exceeds substantially all of the fair value of the underlying leased asset.
5. **Asset specificity test.** The underlying leased asset is of a specialized nature, such that the leased asset is expected to have no alternative use to the lessor at the end of the lease term.

Again, if any of the five criteria are met, then a lessee would classify a lease as a finance lease. The predecessor lease accounting standards used similar lease classification criteria. However, some changes have been made to the updated classification criteria contained in ASC 842. First, the lease term test and the lease payments test were previously based on bright-line tests. The bright-line tests are now principles-based approaches, as described above. In particular, under the predecessor accounting standards for leases, if a lease term was greater than or equal to 75% of the asset's remaining economic life, the lessee would classify the lease as a finance lease. The new standards no longer use this bright-line test, although the 75% threshold could be used as evidence that the lease term is for the major part of the remaining economic life of the underlying leased asset. In addition, if the present value of the sum of the lease payments and any residual value guaranteed by the lessee was greater than or equal to 90% of the asset's fair value, the lessee would have classified the lease as a capital lease (now called a finance lease) under the predecessor accounting standard. Again, the current rules allow for the 90% threshold as evidence of the lease meeting the lease payments test, but the new standards no longer use the bright-line test as the sole determining factor.

Second, ASC 842 includes a fifth criterion that was not included in the predecessor lease accounting guidance related to the specialized nature of the asset. This additional criterion makes it more difficult to classify a lease as an operating lease. If the underlying leased asset is of a highly specialized nature, such an asset would be prohibitively expensive for the lessor to repurpose for use by another user and would have no alternative use to the lessor. In this case, the lessee would classify the lease as a finance lease (the lessor would generally classify the lease as a sales-type lease, as noted in the section discussing lessor accounting).

Note that the dual classification model primarily impacts how a lessee will report lease expense in its income statement subsequent to the initial recognition of the lease. The distinction between operating and finance leases does not impact the recognition of lease-related assets and liabilities on a lessee's balance sheet. Lastly, there is an exemption that applies to the general requirement that lessees recognize the lease-related assets and liabilities on their balance sheets. In particular, when a lease arrangement has a term of 12 months or less, and it does not include an option to purchase the underlying leased asset that the lessee is reasonably certain to exercise, the lessee does not need to report the lease-related asset and liability on its balance sheet.

Lessors must follow the same five fundamental criteria used by lessees to determine the appropriate classification of a lease. However, under US GAAP,

there also exist additional criteria that apply only to lessors and are used to determine whether a lessor should classify a non-sales-type lease as either an operating lease or a direct financing lease. These additional criteria will be discussed in the section addressing lessor accounting.

86.3.2 Discussion of Important Assumptions, Estimates, and Inputs

The classification, initial recognition and measurement, and subsequent measurement of a lease arrangement require lessees and lessors to make certain assumptions and estimates. For example, accounting for leases requires an estimate of the lease term, an appropriate discount rate to calculate the present value of lease payments and any residual value guarantees, and an estimate of the fair value of the underlying leased asset. These assumptions and estimates are discussed in this section.

86.3.2.1 Lease Term

While the lease term may be explicitly stated in a lease contract, a lessee and lessor must also consider contractual provisions such as renewal and termination options. Only if a renewal or termination option is reasonably certain to be exercised should it be factored into the lease term. ASC 842-10-30-1 considers the lease term to be the noncancellable period of the lease, plus:

1. Lease renewal periods, if the lessee is reasonably certain to exercise an option to extend the lease arrangement.
2. Lease termination periods, if the lessee is reasonably certain not to exercise an option to terminate the lease.
3. Lease renewal and termination periods for which the options to extend the lease arrangement or not to terminate the lease arrangement are controlled by the lessor.

Naturally, the assessment of reasonably certain requires professional judgment based on the facts and circumstances present at the commencement of a lease. Judgment should be based on, for example, historical practices and a lessee's or a lessor's economic incentives to exercise renewal or termination options. The FASB suggests that reasonably certain generally would be considered in the context of a threshold of 75%.

Note also that one of the criteria governing the classification of a lease requires a comparison of the lease term and the remaining economic life of the underlying leased asset. The economic life of an asset is defined in ASC 842 as either the period over which an asset is expected to be economically usable by one or more users or the number of production or similar units expected to be obtained from an asset by one or more users. Estimating the economic life of an asset at lease commencement also requires judgment, similar to estimating the depreciable life of an asset purchased.

86.3.2.2 Lease Payments and Residual Values

The lease payments represent the consideration exchanged by a lessee to obtain the right to control the use of the underlying leased asset. For classification purposes, the lease payments test is satisfied if the present value of the sum of the lease payments and any residual value guaranteed by the lessee (that is not already part of the lease payments) equals or exceeds substantially all of the fair value of the underlying leased asset. As discussed above, and as a general rule of thumb, evidence of satisfying the lease payments test could include situations in which the present value of the sum of the lease payments and any residual value guaranteed by the lessee equals or exceeds 90% of the fair value of the underlying leased asset.

ASC 842-10-30-5 discusses the types of lease payments that should be considered at the commencement date to assess the lease payments criterion for lease classification purposes. These lease payments include the following:

1. Fixed payments, net of any lease incentives that have been or will be paid to the lessee.
2. Variable lease payments (e.g., that depend on a market interest rate), which should initially be measured using the referenced index or rate at the lease commencement date.
3. The exercise price of a purchase option that the lessee is reasonably certain to exercise.
4. Amounts paid as penalties for terminating the lease arrangement if the lessee's option to terminate the lease was reflected in the determination of the lease term.
5. Amounts paid by the lessee to a special-purpose entity to structure the lease arrangement. Importantly, note that these amounts should not be included in the determination of the fair value of the underlying leased asset.
6. Amounts that are probable of being owed under a residual value guarantee (this applies to the lessee only).

The residual value is the estimated fair market value of the leased property at the end of the lease term. The lessor can transfer the risk of loss to the lessee through a guarantee of the estimated residual value. A residual value guarantee is a guarantee made by the lessee to the lessor that the value of the underlying leased asset returned to the lessor will be at least some specified amount at the end of the lease term.¹⁰ A lessor may also obtain insurance for the residual value of the underlying leased asset

¹⁰It is important to recognize that lease classification and lease measurement (e.g., the measurement of the lease liability) are not necessarily the same. For example, the lease payments classification criterion requires a lessee and lessor to consider the present value of the sum of the lease payments and any residual value guaranteed by the lessee. In contrast, for purposes of measuring the lease liability, only lessee residual value guarantees that are *probable* of being paid are considered. In addition, the lease liability reflects the present value of the remaining lease payments, while the lease payments classification criterion considers all payments, including payments made before the commencement date.

from a third party (rather than from the lessee), which would not be considered by the lessee in terms of lease classification or lease measurement.

86.3.2.3 Discount Rate

Lessees and lessors need to determine an appropriate discount rate for lease classification and lease measurement purposes. ASC 842 indicates that a lessor should use the rate implicit in the lease. The basic idea is that the rate implicit in the lease is the rate of interest that sets the present value of the lease payments and the residual value of the underlying leased asset (whether guaranteed or unguaranteed) equal to the sum of the fair value of the underlying leased asset (net of any investment tax credits that can be used by the lessor) and the lessor's deferred initial direct costs (if any).

Thus, the rate implicit in the lease is effectively the internal rate of return. For instance, suppose the lessee and lessor enter into a lease contract that requires the lessee to make ten annual lease payments of \$3,000 at the beginning of each year (the first payment is due on the lease commencement date). The lessor estimates the fair value of the asset to be \$20,000 and expects that the asset has no residual value at the end of the lease term. In addition, the lessor has no deferred initial direct costs.¹¹ The interest rate implicit in the lease is the value r in the following equation:

$$3,000 + \sum_{t=1}^9 \frac{3,000}{(1+r)^t} = 20,000$$

A lessee should also use the rate implicit in the lease, unless the lessee is unable to determine that rate. A lessee often is unable to determine the rate implicit in the lease, for example, because the lessee may not have access to the same information that a lessor has related to the fair value of the underlying leased asset, the lessor's residual value estimate, and the initial direct costs incurred by the lessor. In this case, the lessee should use its incremental borrowing rate. ASC 842 indicates that a lessee's incremental borrowing rate should be based on a collateralized borrowing rate using otherwise similar terms reflected in the lease arrangement (e.g., similar duration, amount, and economic environment). Thus, ASC 842 requires the incremental borrowing rate to be a secured rate. Unsecured sources of funds can no longer be considered in determining the incremental borrowing rate, as was the case under the predecessor lease accounting standards.

86.3.2.4 Fair Value of the Underlying Leased Asset

In evaluating the lease payments criterion for classification purposes, the present value of the sum of the lease payments and any residual value guaranteed by the lessee must be compared to the fair value of the underlying leased asset. Specifically, note that the lease payments criterion requires that the fair value of the underlying

¹¹ The lessor will defer initial direct costs if, at the lease commencement date, the fair value of the underlying leased asset is equal to its carrying value. In such a case, the deferred initial direct costs would be included in the determination of the rate implicit in the lease.

leased asset be used, not the fair value of the right to use the underlying leased asset for a period that is typically less than the economic life of the asset. Thus, the fair value of the leased asset should reflect the price that would be received to sell the asset in an orderly transaction on the commencement date. ASC 842-30-55-3 indicates that in some cases, it may be impracticable to determine the fair value of the underlying leased asset. If that is the case, lease classification can be determined without regard to the lease payments test.

86.3.3 Accounting and Reporting Requirements

Once the appropriate classification is determined by the lessee and lessor, the transaction can be appropriately measured and reported in their respective financial statements. This section briefly discusses the initial and subsequent accounting for a lease arrangement. This section first focuses on lessee accounting, and then turns to lessor accounting.¹² It is important to understand that the lease classification criteria do not impact a lessee's initial balance sheet recognition of the lease arrangement. However, the classification as an operating versus a finance lease will ultimately impact how the lease expense is recognized in the income statement. In contrast, the lease classification will, to some extent, impact a lessor's recognition of the lease arrangement.

86.3.3.1 Lessee

A lessee measures the lease liability on the commencement date as the present value of the *remaining* lease payments, as discussed above and defined in ASC 842-10-30-5. As noted previously, a lessee should use the rate implicit in the lease (if known) or, alternatively, its incremental borrowing rate in its determination of the present value of the remaining lease payments. Note that the lease liability does not include amounts transferred to the lessor before the commencement date, and it only includes residual value guarantees that are probable of being owed by the lessee.¹³

A lessee measures the right-of-use asset in accordance with ASC 842-20-30-5, which indicates that, at the commencement date, the cost of the right-of-use asset consists of the following:

¹²See, for example, Gordon et al. (2019, Chap. 18) and Kieso et al. (2019, Chap. 21) for detailed illustrations of the recognition and measurement of leases from the perspective of both the lessee and the lessor.

¹³As discussed previously, the calculation of lease payments for the lease payments test included in the classification criteria does not necessarily reflect the lease payments included in the lease liability at the commencement date. The lease payments test includes all payments, including payments made prior to the commencement date and a lessee's residual value guarantee, regardless of whether it is probable that the lessee will owe amounts associated with the guarantee (if any).

1. The lease liability at the commencement date
2. Amounts paid to the lessor at or before the commencement date, which are not reflected in the lease liability, net of any lease incentives received
3. The lessee's initial direct costs

Thus, the lease liability and the right-of-use asset may not be the same at the lease commencement date. Note also that ASC 842 changed the criteria for determining initial direct costs for both the lessee and the lessor. ASC 842 defines initial direct costs as incremental lease-related costs that would not have been incurred if the lease had not been obtained. Under the predecessor lease accounting guidance, these incremental costs included items arising before the lease was executed (such as legal fees) that generally are not included as initial direct costs under ASC 842. An example of an initial direct cost that would be included is lease documentation costs such as filing fees. Excluded costs would include items such as advertising, lease negotiation fees, and employee salaries. Specifically, any costs incurred that would have been incurred even if the lease arrangement ultimately failed to materialize would not be included as initial direct costs.

Following the initial measurement of the lease liability and the right-of-use asset, the lessee will make periodic lease payments while amortizing the right-of-use asset and recording interest expense associated with the lease liability. Importantly, the pattern of recognizing amortization and interest expense will differ depending on how the lessee classifies the lease – that is, as either a finance lease or an operating lease.

With a finance lease, the lessee will simply amortize the right-of-use asset on a straight-line basis, generally over the lease term.¹⁴ Interest expense related to the lease liability is determined using the effective interest method (and the same discount rate used to measure the lease liability on the commencement date). As a result, total lease expense typically declines over the course of the lease term – while amortization of the right-of-use asset is constant, interest expense declines as the periodic lease payments reduce the lease liability.

With an operating lease, the annual lease expense is recognized on a straight-line basis and is reflected as a single expense in the income statement and a single item in the operating section of the statement of cash flows. The annual lease expense impounds both interest and amortization expense. The annual lease expense is generally the sum of the total undiscounted cash flows (total lease payments plus initial direct costs minus lease incentives) divided by the lease term. Each period the lessee calculates interest expense (or the portion of the annual lease expense related to interest) using the effective interest method, similar to the approach used for a finance lease. However, the annual amortization of the right-of-use asset is now the difference between the annual lease expense and the portion of the annual lease

¹⁴Note, for example, that if the lease contains a purchase option that the lessee is reasonably certain to exercise, the right-of-use asset would be amortized over the useful life of the asset rather than the lease term.

expense related to interest for the period. As a result, amortization of the right-of-use asset will be higher at the end of the lease term when interest expense is lower. It is important to reiterate that the lessee is not reporting interest expense and amortization expense. Instead, the lessee is reporting a single lease expense that is constant each period. Included in the annual lease expense are interest and amortization. The calculation of interest is simply used to determine the annual accretion to the lease liability and the amortization of the right-of-use asset.

86.3.3.2 Lessor

The accounting rules related to the lessor were generally not impacted by the adoption of ASC 842. Similar to the lessee, a lessor follows the same fundamental criteria contained in ASC 842-10-25-2 (which were discussed in Sect. 86.3.1) when determining the appropriate classification of a lease. If any of the criteria described in ASC 842-10-25-2 are met, the lessor will classify the lease as a sales-type lease, indicating that the lease transfers control of the underlying leased asset to the lessee.

For a sales-type lease, a lessor will derecognize the underlying leased asset and record a financial receivable reflecting the lessor's net investment in the lease at the commencement date. The net investment in the lease reflects the lease receivable and the unguaranteed residual value of the underlying leased asset. The lease receivable is the present value of the remaining lease payments and the guaranteed residual value (whether guaranteed by the lessee or a third party). The unguaranteed residual value is based upon the amount the lessor expects to derive from the underlying leased asset at the end of the lease term that is not guaranteed by the lessee or any third party. As discussed before, these amounts must be discounted using the lessor's implicit rate in the lease. Initial direct costs are expensed at the commencement date unless there is no profit or loss on the lease (i.e., the fair value of the leased asset is equal to its carrying amount), in which case the lessor's initial direct costs are deferred and recognized over the lease term.

ASC 842 indicates that the selling profit or loss on a sales-type lease at the commencement date is equal to the lower of (1) the fair value of the underlying leased asset or (2) the lease receivable plus any lease payments made by the lessee before the commencement date, minus the following:

- (a) The carrying amount of the underlying leased asset net of any unguaranteed residual value related to the underlying leased asset
- (b) The lessor's deferred initial direct costs (if any)

The sales price (revenue) is generally equal to the fair value of the underlying leased asset; however, it could be lower in some circumstances, such as if there is an unguaranteed residual asset. The present value of an unguaranteed residual asset also reduces cost of goods sold so that it is below the asset's cost. Following the initial recognition of a sales-type lease, a lessor will recognize interest income on the net investment in the lease using the effective interest method while collecting lease payments from the lessee (which reduce the lessor's net investment in the lease).

If the ASC 842-10-25-2 criteria are not met, the lessor considers additional criteria described in ASC 842-10-25-3, which determines whether the lessor will classify the lease as an operating lease (like the lessee) or, alternatively, a direct financing lease. Specifically, a lessor will classify the lease as an operating lease unless both of the two following criteria are met (after failing to meet the basic criteria in ASC 842-10-25-2):

1. **Expanded lease payments test.** The present value of the sum of the lease payments and any residual value guaranteed by the lessee (that is not already part of the lease payments) and/or any residual value guaranteed by a third party unrelated to the lessor equals or exceeds substantially all of the fair value of the underlying leased asset.
2. **Collectability test.** The lease payments plus any amounts needed to satisfy a residual value guarantee are probable of being collected by the lessor.

A key feature of the above criteria is that the lessor now includes third-party residual value guarantees in the lease payments test. If the lease payments test is met following the inclusion of third-party residual value guarantees, and payments are probable of being collected, then the lessor will classify the lease as a direct financing lease. Otherwise, the lessor will use the operating lease classification, similar to the lessee.

The accounting for direct financing leases and operating leases is different. At initial recognition for a direct financing lease, a lessor recognizes its net investment in the lease and derecognizes the underlying leased asset, similar to the initial recognition of a sales-type lease. However, unlike a sales-type lease, the lessor of a direct financing lease will include any selling profit and initial direct costs in the net investment in the lease to be recognized over time.¹⁵ To be consistent with the FASB's revenue guidance, ASC 842 precludes the recognition of selling profit at lease commencement for a direct financing lease because the lease does not transfer control of the underlying leased asset. Instead, selling profit is recognized as interest income over the lease term, consistent with the notion that a direct financing lease effectively converts the lessor's risk related to the underlying leased asset into credit risk. Note that a critical difference between a sales-type lease and a direct financing lease is not the amount of income recognized on the net investment in the lease, but rather the timing of income recognition. Another, perhaps less substantial, difference is the timing of the recognition of initial direct costs.

Finally, with an operating lease, the lessor's asset related to the lease remains on the lessor's balance sheet and will be depreciated by the lessor while the lessor recognizes rent revenue associated with the lease. No lease-related asset or liability will be recognized at lease commencement.

In general, a lessee and lessor should arrive at the same conclusion with respect to lease classification outcomes. For example, if a lessee determines that a lease is

¹⁵ A selling loss at the commencement date would be recognized immediately.

appropriately classified as a finance lease, the lessor should generally conclude that the lease is a sales-type lease. However, because of differences in inputs used in determining the classification of a lease, the lessee and lessor could arrive at different conclusions. For example, the lessor may use the rate implicit in the lease while the lessee uses its incremental borrowing rate (see Sect. 86.3.2.3), or the lessee and lessor may arrive at different conclusions with respect to the likelihood of a renewal option being exercised. In addition, a lessee and lessor may not both classify a lease as an operating lease because a lessor will also consider arrangements unrelated to the lessee, such as a residual value guarantee obtained from a third party.

86.3.4 Further Discussion of Lease Arrangements

There are various general forms that a lease arrangement can take. We can think of these general lease forms as (1) direct leases, (2) leveraged leases, and (3) sale and leasebacks. So far, we have generally assumed that a lease is a direct lease, whereby the lessee obtains the right to control the use of a lessor's identified asset for a specified period of time in exchange for consideration paid to the lessor. This relatively simple arrangement may not always be the case.

For example, one type of lease arrangement that was once commonly used is a leveraged lease, as determined by the criteria contained in the predecessor leasing standard – specifically, ASC 840-10-25-43(c). The leveraged lease classification applied only to lessors – the lessee would continue to classify the lease as either an operating lease or a finance lease (previously referred to as a capital lease). From the lessor's perspective, leveraged leases are direct financing leases that involve a lessee, a lessor, and a long-term creditor. ASC 840 required that the financing provided by the long-term creditor (for the lessor to acquire the asset) be nonrecourse as to the general credit of the lessor (although the creditor could have recourse to the specific property leased and the unremitted rentals relating to it). The amount of financing needed to be sufficient to provide the lessor with substantial leverage in the transaction. Such leases were typically desirable for the tax benefits obtained by the lessor (as long as the lessor had a sufficient equity stake in the asset) that typically also reduced the cost of the lease to the lessee. However, tax law changes reduced the desirability of leveraged leases such that few such transactions have been entered into in recent years. In addition, ASC 842 no longer contains guidance for leveraged leases and, thus, does not permit new leases to be accounted for as leveraged leases.

Finally, an arrangement that is frequently used today is known as a sale and leaseback. In a sale and leaseback arrangement, a company (the seller-lessee) sells an asset that it owns to a second party (the buyer-lessor), and the two parties then enter into a lease in which the seller-lessee leases back the asset from the buyer-lessor for some or all of the asset's remaining economic life. The buyer-lessor retains title to the asset and all other benefits of ownership, such as tax credits and depreciation allowance. The seller-lessee receives funds from the sale of the asset along with the right to use the asset and, in exchange, makes periodic lease payments to the buyer-lessor.

Although sale and leaseback accounting applied only to lessees under ASC 840, both lessees and lessors must apply sale and leaseback accounting under ASC 842. In order to qualify for sale and leaseback accounting, first, the transaction has to qualify as a sale in accordance with the FASB's revenue accounting standard. If the transaction qualifies as a sale, then the seller-lessee and buyer-lessor will separately account for the sale (in accordance with the revenue standard) and the leaseback (in accordance with the leasing standard). If the transaction qualifies as a sale, but the seller-lessee classifies the lease as a finance lease and the buyer-lessor classifies the lease as a sales-type lease, the arrangement would be considered a "failed" sale and leaseback. Control of the asset is effectively transferred from the seller-lessee to the buyer-lessor and then back to the seller-lessee. As a result, the guidance in ASC 842-40-25-5 requires a failed sale and leaseback to be accounted for as a financing event by the seller-lessee and as a lending event by the buyer-lessor.

86.3.5 Important Differences between US GAAP and IFRS

While the lease accounting rules developed as part of the joint project between the FASB and International Accounting Standards Board (IASB) have many similarities, the boards came to different conclusions in some aspects. One of the most substantial differences is the fact that the FASB retained its dual model for lessees (operating leases versus finance leases) while the IASB chose to require that lessees apply a single model to all leases. Under the new IFRS standards for leases (i.e., IFRS 16, *Leases*), the lessee does not differentiate between operating leases and finance leases (IASB 2016). Instead, under IFRS 16, lessees treat all leases similar to how a finance lease would be treated for US GAAP purposes. As a result of the dual vs. single model, differences arise not only in terms of lease classification but also in the resulting income statement and statement of cash flow effects due to the fact that, for lessee operating leases under US GAAP, only a single lease expense amount is recognized (on a straight-line basis) with the cash outflow reported as an operating activity in the statement of cash flows.

Although lessees use a single lease model under IFRS, IFRS 16 does contain the lease classification criteria included in ASC 842. However, for IFRS purposes, the lease classification criteria are intended solely for the lessor. Some relatively minor differences exist for lessor accounting. Under US GAAP and IFRS, a lessor may recognize profit on a sales-type lease. However, under IFRS, a lessor can also recognize profit on a direct financing lease if the performance obligations have been met (this is not permitted under US GAAP, as discussed in the section that discusses [lessor](#) accounting).

Like the new US GAAP standards, IFRS also maintains an exemption for reporting leases on the balance sheet if the lease term is less than 12 months in duration. Additionally, under the new IFRS standards, a lessee may elect to treat leased assets with a value of less than \$5,000 as rental agreements. For example, if a

lessee leased a car with a value of \$3,000 and a useful life of 2 years, the lessee does not have to recognize a lease asset or a lease liability on its balance sheet.

86.3.6 Impact of New Leasing Accounting Guidance

As discussed above, ASC 842 was issued in 2016 and effective for publicly traded companies during a company's first fiscal period beginning after December 15, 2018. This means that most companies first implemented the new leasing guidance for their fiscal years beginning on January 1, 2019. As such, researchers have only just begun to explore the effects of the new lease accounting guidance.

The new leasing standards are expected to bring a greater degree of transparency to financial statements than previously available and are expected to significantly impact financial statements and important financial ratios for many companies. For instance, the updated guidance under both US and international accounting standards should enable analysts and investors to compare the credit capacity and leverage metrics of companies with different financing practices much more readily and consistently by eliminating the off-balance-sheet treatment of operating leases. This will impact reported leverage for many companies, as recognizing an asset and liability from leases will cause leverage metrics to increase. The recognition of amounts related to operating leases that were once only disclosed in the notes to the financial statements could impact how investors and lenders perceive the company.

Many questions remain unanswered since the FASB's and IASB's leasing standards became effective. Do the new lease accounting requirements provide decision-useful information to investors and creditors? Have the new leasing standards impacted the comparability of financial information? Have the new leasing standards impacted companies' operating and financing decisions? Has the use or structuring of leases changed as a result of the new guidance? What are the consequences of the FASB's dual classification approach in comparison to the IASB's single model approach? In sum, it remains to be determined whether the standard setters' objectives of increasing the amount of lease-related decision-useful information reported by lessees and lessors have been met, as well as whether there have been any unintended consequences associated with the adoption of ASC 842 and IFRS 16.

86.4 Conclusion

Capital investment decisions are among the most critical decisions made by corporate managers. When confronted with the decision to make capital investments, managers must identify a viable source of financing. Theory suggests that under certain conditions, the lease-versus-buy decision should be irrelevant to a company. However, prior research has explored and provided evidence of the factors contributing to the widespread use of leases.

This chapter focuses on companies' use of leases as a financing mechanism to facilitate capital investments. We discuss the nature and use of leases, as well as the economic theory rationalizing the existence and implementation of lease arrangements. We then turn our attention to recent changes to US and international accounting standards that have ended several decades of off-balance-sheet reporting for many lease transactions. We discuss how the recent changes in accounting and financial reporting requirements have potentially shifted the expected costs and benefits of some lease arrangements. We then conclude by highlighting the need for researchers to study the impact of the new lease accounting standards that have recently been adopted in countries around the globe.

References

- Altamuro, J., R. Johnston, S. Pandit, and H. Zhang. 2014. Operating leases and credit assessments. *Contemporary Accounting Research* 31 (2): 551–580.
- Ang, J., and P.P. Peterson. 1984. The leasing puzzle. *The Journal of Finance* 39 (4): 1055–1065.
- Beatty, A., S. Liao, and J. Weber. 2010. Financial reporting quality, private information, monitoring, and the lease-versus-buy decision. *Journal of Accounting Research* 85 (4): 1215–1238.
- Bratten, B., P. Choudhary, and K. Schipper. 2013. Evidence that market participants assess recognized and disclosed items similarly when reliability is not an issue. *The Accounting Review* 88 (4): 1179–1210.
- Caskey, J., and N.B. Ozel. 2019. Reporting and non-reporting incentives in leasing. *The Accounting Review* 94 (6): 137–164.
- Cornaggia, K.J., L.A. Franzen, and T.T. Simin. 2013. Bringing leased assets onto the balance sheet. *Journal of Corporate Finance* 22: 345–360.
- Dhaliwal, D., H.S. Lee, and M. Neamtiu. 2011. The impact of operating leases on firm financial and operating risk. *Journal of Accounting, Auditing & Finance* 26 (2): 151–197.
- Eisfeldt, A.L., and A.A. Rampini. 2009. Leasing, ability to repossess, and debt capacity. *The Review of Financial Studies* 22 (4): 1621–1657.
- Ellickson, R.C., and C.D. Thorland. 1995. Ancient land law: Mesopotamia, Egypt, Israel. *Chicago-Kent Law Review* 71 (1): 321–411.
- Ely, K. 1995. Operating lease accounting and the market's assessment of equity risk. *Journal of Accounting Research* 33 (2): 397–415.
- Financial Accounting Standards Board (FASB). 1976. *Statement of Financial Accounting Standards No. 13. Accounting for leases*. Norwalk, CT: FASB.
- . 2016. *Accounting Standards Update No. 2016-02, Leases (Topic 842)*. Norwalk, CT: FASB.
- Gavazza, A. 2010. Asset liquidity and financial contracts: Evidence from aircraft leases. *Journal of Financial Economics* 95 (1): 62–84.
- . 2011. Leasing and secondary markets: Theory and evidence from commercial aircraft. *Journal of Political Economy* 119 (2): 325–377.
- Gordon, M. 1974. A general solution to the buy-or-lease decision. *The Journal of Finance* 29 (1): 245–250.
- Gordon, E.A., J.S. Raedy, and A.J. Sannella. 2019. *Intermediate accounting*. 2nd ed. Hoboken, NJ: Pearson Education, Inc.
- Graham, J., M. Lemmon, and J. Schallheim. 1998. Debt, leases, taxes, and the endogeneity of corporate tax status. *The Journal of Finance* 53 (1): 131–162.
- Imhoff, E.A., R.C. Lipe, and D.W. Wright. 1991. Operating leases: Impact of constructive capitalization. *Accounting Horizons* 5 (1): 51–63.

- . 1997. Operating leases: Income effects of constructive capitalization. *Accounting Horizons* 11 (2): 12–32.
- Imhoff, E.A., and J.K. Thomas. 1988. Economic consequences of accounting standards: The lease disclosure rule change. *Journal of Accounting and Economics* 10 (4): 277–310.
- International Accounting Standards Board (IASB). 2016. *International financial reporting standard 16, leases*. London, UK: IASB.
- Johnson, R., and W. Lewellen. 1972. Analysis of the lease vs. buy decision. *The Journal of Finance* 27 (4): 815–823.
- Kieso, D.E., J.J. Weygandt, and T.D. Warfield. 2019. *Intermediate accounting*. 17th ed. Hoboken, NJ: John Wiley & Sons, Inc.
- Kraft, P. 2015. Rating agency adjustments to GAAP financial statements and their effect on ratings and credit spreads. *The Accounting Review* 90 (2): 641–674.
- Miller, M.H., and C.W. Upton. 1976. Leasing, buying, and the cost of capital services. *The Journal of Finance* 31 (3): 761–786.
- Modigliani, F., and M.H. Miller. 1958. The cost of capital, corporation finance and the theory of investment. *American Economic Review* 48 (3): 261–297.
- . 1963. Corporate income taxes and cost of capital: A correction. *American Economic Review* 53 (3): 433–443.
- Monson, D.W. 2001. The conceptual framework and accounting for leases. *Accounting Horizons* 15 (3): 275–287.
- Myers, S.C., D.A. Dill, and A.J. Bautista. 1976. Valuation of financial lease contracts. *The Journal of Finance* 31 (3): 799–819.
- Rapoport, M. 2013. Criticism mounts against new accounting rules for leases. *The Wall Street Journal* (23 September 2013).
- Securities and Exchange Commission (SEC). 2005. *Report and recommendations pursuant to section 401(c) of the Sarbanes-Oxley act of 2002 on arrangements with off-balance sheet implications, special purpose entities, and transparency of filings by issuers*. Washington, DC: U.S. Securities and Exchange Commission.
- Sharpe, S., and H. Nguyen. 1995. Capital markets imperfections and the incentive to lease. *Journal of Financial Economics* 39 (2/3): 271–294.
- Smith, C.W., and L.M. Wakeman. 1985. Determinants of corporate leasing policy. *The Journal of Finance* 40 (3): 895–908.
- Spencer, A.W., and T.Z. Webb. 2015. Leases: A review of contemporary academic literature relating to lessees. *The Accounting Review* 29 (4): 997–1023.
- Wilkins, T., and I. Zimmer. 1983. The effect of leasing and different methods of accounting for leases on credit evaluations. *The Accounting Review* 58 (4): 749–764.



Pension Accounting, Inside Debt, and Capital Structure

87

Cheng-Few Lee and James Juichia Lin

Abstract

This study aims to examine how pension accounting relates to inside debt and capital structure. We introduce basic concepts of pension plans and pension accounting (including the determinants of pension liabilities, assets, and pension expense reported in financial statement) in accordance with U.S. GAAP (FAS 158) and IFRS (IAS 19). We also provide a review of empirical research on the pension accounting associated with earnings management and firm value. Finally, following Alderson and Lee (J Econ Bus, 40:209–228, 1988), this chapter introduces a developed model in which individuals and corporations have an incentive to engage in pension plans to deals with corporate pension funding policy as part of the capital structure decision.

Keywords

Pension plans · Pension actuarial assumptions · Earnings management · Value relevance · Capital structure decision

JEL Classification

J32 · G32 · M41

Abbreviations

DB Defined benefit

C.-F. Lee

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: lee@business.rutgers.edu; cflee@business.rutgers.edu

J. J. Lin (✉)

Department of Information Management and Finance, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

e-mail: Jameslin@nycu.edu.tw

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_89

2041

87.1 Introduction

Defined benefit (DB) and defined contribution (DC) plans are the two most common types of pension plans. In a DB plan, the employer guarantees to pay the employee a specific level of benefits for life upon the employee's retirement; whereas in a DC plan (including Individual Retirement Accounts (IRAs) and 401 (k) plans), the employer makes a fixed sum of contribution every year to an employee's account but does not guarantee the payout upon the employee's retirement. That is, only the employer's contribution is guaranteed under a DC plan but not the future benefit payouts. In the USA, the maximum annual retirement benefit allowed under a DB plan is \$195,000 whereas the maximum contribution allowed under a DC plan is \$49,000 in 2010 (Calhoun 2010). At an aggregate level, DC plans are cheaper and involve lower risk for sponsoring firms than DB plans. According to a study by PricewaterhouseCoopers (2005), defined benefit plans cost about 5–6% of payroll while defined contribution plans cost only about 3%. Thus, all these studies suggest that DB plans are more costly to the sponsoring firms (Choy et al. 2014).

In addition to the difference in costs, these two types of plans impose different levels of risk on the sponsors. Since sponsoring firms guarantee the payout of a specific number of benefits for the life of their employees upon their retirement under DB plans, firms with DB plans have to bear various uncertainties and risks associated with the future payouts, such as the market and longevity risks (Munnell et al. 2007; Lin and Werner 2021). DB firms face such risks as fluctuations in the assumption regarding the longevity of employees, interest rate, inflation, and rate of investment return. Any change in these factors can have a significant effect on the required payout or the funds available for distribution, and enhance the contribution required in the current period to meet these expected future payouts.

This study aims to examine how pension accounting relates to inside debt and capital structure. The remainder of this chapter is organized as follows. The next section describes pension plan freeze and termination. Section 87.3 discusses the effects of ERISA and PBGC on Pension Management. Section 87.4 presents accounting for defined pension plans under FAS 158 and IAS 19. Sections 87.5 and 87.6 provide a review of empirical research on the pension accounting associated with earnings management and firm value. Section 87.7 introduces a developed model in which individuals and corporations have an incentive to engage in pension plans to deals with corporate pension funding policy as part of the capital structure decision.

87.2 Pension Plan Freeze and Termination

Traditionally, the occupational pension system was designed and dominated by the DB plans while DC plans were only offered by small firms or as supplementary plans. However, there has been a steady increase in the adoption of DC plans since

1975, the year following the passage of the Employee Retirement Income Security Act (ERISA).¹ In the last two decades, a shift in pension coverage from defined benefit (DB) plans to defined contribution (DC) plans has been underway in the United States and UK (Munnell et al. 2007). According to the 2008 Private Pension Plan Bulletin report, based on the information collected from US firms' Form 5500 filings with the U.S. Department of Labor, the number of DB pension plans decreases by 42.3%, from 20,035 in 1975 to 11,557 in 2005. On the contrary, the number of DC plans multiplies by about 7.8-times from 8,587 in 1975 to 67,678 in 2005.²

During the 1980s and early 1990s, DB pension freeze or termination is limited mostly to firms facing bankruptcy or engaged in merger and acquisition activities (Choy et al. 2014). However, the shift in pension coverage from DB to DC plans has gained popularity among large healthy firms since the early 2000s (Munnell et al. 2007). Watson Wyatt (2010) documents an increase in DB plan freezes among Fortune 100 firms. The number of DB plan freezes rose from 45 in 2004 to 113 in 2006. McFarland et al. (2009) observe 82 US public companies that froze their DB plans in the period 2003–2007. Some firms that froze their plans during the period include Motorola Inc. that froze its pension plan to new employees in 2004; Lockheed Martin, Hewlett-Packard, and Sears Holdings followed suit and froze their DB plan in 2005. In 2006, Coca-Cola Bottling and IBM both decide to impose a total freeze on their DB plans in 2006 (Munnell et al. 2007). These large healthy companies either offer a 401-K plan or improve their existing 401-K plans after the freeze.

87.3 Effects of ERISA and PBGC on Pension Management

Employee Retirement Income Security Act (ERISA) of 1974 improves the protection of retirement benefits for private-sector employees. Two provisions of ERISA affect the firm's decision to terminate the plan or to exercise the pension plan put option. The first provision establishes a rule for determining the minimum benefit level to which employees are entitled. When a pension plan is underfunded, the put option has a positive value. The employee will receive the smaller of the guaranteed vested benefits or the fund's value. A second ERISA provision establishes pension plan termination insurance under the Pension Benefit Guaranty Corporation (PBGC). Assuming a pension fund's value is so low on termination that an addition of 30% of the firm's equity and the guaranteed pension liabilities. This limits the maximum exposure of the firm to 30% of its equity for meeting underfunded pension

¹Source: U.S. Development of Labor Employee Benefits Security Administration, Private Pension Plan Bulletin, February 2008.

²Only pension plans with 100 or more participants are included in the statistics. The statistics include single employer plans, plans of controlled groups of corporations, and multiple-employer non-collectively bargained plans.

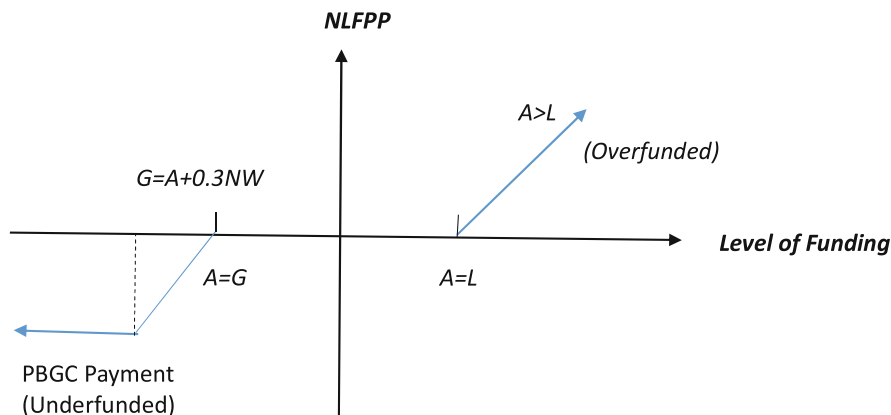


Fig. 1 Net liability of firm in pension plan (NLFPP)

liabilities. In exchange for this protection, the value pays a fixed premium based on a per-employee rate.

The effects of the PBGC and the pension fund assets on firm value are shown in Fig. 1. When the value of the pension plan's assets (A) exceeds the accrued liabilities (L), the retirees will receive all of their benefits and the firm has the right to $(A-L)$, or the amount of overfunding of the pension plan. If the value of the plan's asset (A) equals the accrued benefits (L), or $(A-L)$ equals 0, the retirees will receive their guaranteed benefits and the firm has no interest in the pension fund. If $(A-G)$ is less than 0, where G is the level of guaranteed benefits, the firm must make up the difference so that the retirees receive their guaranteed benefits, but only up to assumed 30% of the firm's net worth (NW). The additional funding to meet the shortfall will come from the PBGC when $G > A + 0.3NW$.

87.4 Accounting for Pension Plans Under FAS 158 and IAS 19

Following fair-value pension accounting, U.S. GAAP (FAS 158) and IFRS (IAS 19) recognize the difference between pension asset and pension liability (defined benefit obligation, DBO) as net pension asset or on liability referred to as funded status reported in the statement of financial position. Pension assets are employer contributions and the actual return on plan assets at fair value which increase pension plan assets. Employer's defined pension obligation (DBO or PBO) is the present value of the deferred compensation obligation for vested and non-vested employees at future salaries level under the terms of the pension plan. Actuarial assumptions (i.e., discount rate, salary growth rate, and longevity) are responsible for determining the pension obligation.

Following FAS 158 or IAS 19, all changes in the defined benefit obligation and plan assets in the current period be recognized in comprehensive income (including net income and other comprehensive income). The following items account for the

changes in the defined benefit obligation (Kieso, et al. 2020). (1) service cost: referred to as the increase in DBO from employee service in the current period. (2) past service cost: referred to the change in the DBO resulting from a plan amendment. (3) accrued interest expense based on the yields of high-quality bonds. (4) benefits paid to retired employees (beneficiary). (5) gains or losses related to the DBO reported in the section of other comprehensive income, net of tax. The items that account for the change in pension assets are as follows. (1) employer contributions to plan assets. (2) actual return on the plan assets. (3) benefits paid to retired employees (beneficiary). Pension expense reported in the income statement consists of the service cost and net interest expense. Net interest expense is the difference between the accrued interest expense from PBO and the expected rate of return on pension assets. The difference between the actual return and the expected return is recognized as other comprehensive income³.

87.5 Pension Assumptions and Earning Management

Pension accounting which deals with recognition and valuation of pension-related liabilities, assets, and expense, has been developed for decades but debates on its flexible discretion and information content still remain. If firms intentionally obscure the economic status of DB plans by using pension accounting measurement and recognition guidance to manage the financial statements, inaccurate valuations may result (Hsieh et al. 2021). According to the Statement of Financial Accounting Standards 158, hereafter SFAS 158, the actuarial parameters used to compute projected benefit obligations (such as assumed discount rate) and the periodic pension costs should reflect the rate at which the pension benefits could be effectively settled. Although the standard requires the best estimates, it does allow certain discretion in setting these actuarial pension assumptions. For example, in setting the discount rate, the plan manager could refer to the rates of return on high-quality fixed-income investments with the same maturity as the pension obligation, either currently available or expected to be available in the future. This practice provides management with the discretion to choose a desirable discount rate. Further, relatively small changes in assumed pension rates can have a significant impact on corporate pension value, reported earnings, and firm value (Glaum 2009).² Regulators have also expressed some concern about this issue. The U.S. Securities and Exchange Commission (SEC) has instituted a review process for certain filings with expected rates of return on pension plan assets in excess of 9% (Dale 2003). In

³The major differences between FAS 158 and IAS 19 are as follows. (1) FAS 158 amortizes PSC over the remaining service lives of employees, while IAS 19 recognizes past service cost as a component of pension expense in income immediately. (2) FAS 158 recognizes liability and asset gains and losses in “Accumulated other comprehensive income” and amortizes these amounts to income over remaining service lives. Under IAS 19, companies recognize both liability and asset gains and losses (referred to as re-measurements) in other comprehensive income. (Kieso et al. 2020)

addition, by SEC mandate, the Financial Accounting Standards Board (FASB) incorporated an overhaul of pension accounting that was carried out jointly with the International Accounting Standards Board (IASB).³

In addition to the discount rate, the management can also exercise discretion in choosing the expected salary growth rate and the expected rate of return on plan assets. The expected salary growth rate is used to estimate the future pay level of the plan's active participants, which in turn affects the sponsor's pension obligation. The expected rate of return (ERR) on the pension assets' assumption captures the expected future cash inflow from pension assets. A higher ERR leads to a lower sponsor's current obligation. As ERR is typically determined by managers at the beginning of the fiscal year, the manager evaluates the current plan conditions and the asset plan operating prospects for the current year (Picconi 2006). By altering the pension plan assumptions, the sponsor's management can influence its obligations and the required contribution to the plan (Hsu et al. 2013; Lin and Werner 2021).

The potential impact of DB plan reporting improprieties on company valuations by the capital markets evokes the consideration of factors that might influence managers to intentionally manipulate pension assumptions under pension accounting guidance as a means to manage overall accounting reports. Much of the evidence in this regard reveals that pension accounting assumptions are highly correlated with the achievement of specific pension-related goals as well as larger business activity-related objectives. For instance, Gopalakrishnan and Sugrue (1995) find that leverage and pension plan funding play important roles in the choice of these assumptions, while Godwin et al. (1996) find that managers are likely to increase the interest rate assumption in response to declines in earnings, increasingly restrictive dividend constraints, and tightening debt covenants. Bergstresser et al. (2006) document that managers are more aggressive with the estimated ERR in pension obligation computation before undertaking acquisitions when their earnings are near critical thresholds, and before they exercise stock options. They, however, lower the expected rate assumption to reduce income before CEO option grants. Further, Comprix and Muller (2006) find that managers strategically increase the ERR estimation to boost pension income before the CEO's open market stock sales. Comprix and Muller (2011) suggest that employers bias their accounting assumptions downward to exaggerate their pension burden when they want to freeze their pension plans. Amir and Gordon (1996) find that when a firm is close to violating the debt covenant, the management selects a discount rate, and those health care cost assumptions that understate the post-retirement benefit obligations.

87.6 Pension Accounting Information and Value Relevance

Value relevance studies, also called association studies, are one way of examining whether financial statement information is decisional useful by the explained proportion of variation in stock prices. Landsman (1986) makes the first study to examine the value relevance of pension information. He concludes that pension information which has been disclosed under FAS 36 at the time is value relevant.

Landsman and Ohlson (1990) examine the extent to which stock price fully reflects the off-balance-sheet pension item (net pension liability). They document that the market underprices pension disclosures. Barth et al. (1993) find that the fair value of pension asset and PBO disclosed in the pension note under SFAS 87 are associated with stock price. Coronado and Sharpe (2003) follow a similar research design and show that market tends to pay more attention to the recognized pension information in balance sheet and income statement than to disclosed information. Davis-Friday et al. (1999) also document that recognized post-retirement benefits other than pension liabilities receive greater weight than when those same items are disclosed only. Wiedman and Wier (2004) evidence that investors view under-funded pension plans as liabilities to the sponsoring firm, but do not view over-funded plans as assets to the firm. In addition, Picconi (2006) examines the ability of investors and analysts to fully assimilate available pension information. He finds that neither prices nor forecasts fully reflect the quantifiable future earnings effects of changes in pension information at the time it is released in the firm's 10-K report. Franzoni and Martin (2006) argue that the market significantly overvalues firms with severely underfunded pension plans. The evidence shows that those firms continue to earn lower stock returns than firms with healthier pension plans for at least 5 years after the first emergence of the underfunding. Besides, Hann et al. (2007a) find that the discretionary component of PBO is incrementally value-relevant beyond the non-discretionary component ones.

Accordingly, the examinations and comparisons of value relevance under different accounting standards are also documented in the literature. Coronado et al. (2008) argue that more pertinent information on pension finance is relegated to footnotes, which may not receive much attention from investors and portfolio managers. They suggest that the FASB's reform efforts on pension accounting standards should substantially aid the market's ability to value firms with defined benefit pensions. On the contrary, Hann et al. (2007b) compare the value and credit relevance of financial statements under fair-value and smoothing (U.S.'s SFAS 87) models of pension accounting. The result shows that there are no informational benefits of adopting fair-value pension accounting model.

87.7 Pension Plans and Capital Structure Decisions⁴

Inside debts (e.g., pension and other deferred compensation plans) resemble risky corporate debts in that their values are sensitive to both credit risk and liquidation value of corporate assets in bankruptcy (Sundaram and Yermack 2007). Extant literature widely documents that inside debts motivate executives to lower firm risk in order to alleviate creditors' concern over risk-shifting and asset substitution

⁴The following section is original discussed in the paper titled "Corporate Pension Policy and Capital Structure Decisions" published in *Journal of Economics and Business* by Alderson, M. J. and Lee, C. F. (1988).

problems (e.g., Wei and Yermack 2011; Cassell et al. 2012; Wang and Zhang 2014; Chi et al. 2017; Colonnello et al. 2017; Freund et al. 2018). In recent studies, Borah et al. (2020) and Lee et al. (2021) find that firms paying their CEOs with large inside debt leads to reductions in firms' cost of debt, default risk, systematic risk, and idiosyncratic risk.

Following Amott and Gersovitz (1980), this section deals with corporate pension funding policy as part of the capital structure decision. A model is developed in which individuals and corporations have an incentive to engage in contracts for the deferral of wage income in the form of an unfunded pension.

The model assumes that there are two mutually exclusive and exhaustive groups in the economy: capitalists and labor. Capitalists initiate the firm formation and maintain nonneutral investment positions, while labor holds a monopoly on the sale of specialized services. Both groups operate in a one-period, two-date framework. At the beginning of the period (time 0), firms negotiate an agreement with labor, calling for the payment of $\$w$ immediately, in exchange for services to be rendered over the course of the period. Simultaneously, firms issue a series of claims on their state-dependent (end-of-period) operating earnings.

The opportunity to defer wage income from time 0 to time 1 is available. The pension contract is structured such that employees receive their cash wage ($\$w$) at the beginning of the period and immediately return a portion of that wage to the firm in exchange for an unfunded, vested pension claim. From an economic standpoint, the wage deferral arrangement makes the employee a creditor of his/her firm. That relationship serves as the foundation for the integration of corporate pension policy with the capital structure decision.

Both firms and labor are motivated to engage in a pension contract for several reasons. Employees benefit from the deferral of current income, especially if their present marginal tax rate exceeds the rate at which they expect to pay in retirement. Payroll taxes (i.e., FICA) are also a significant source of tax saving because they are levied on current wages and not on pension income. Defined-benefit pension plans also provide nontax advantages to labor by enforcing the discipline to save for retirement and shielding the employee from the investment risk to his/her retirement saving.

The employer firm also recognizes several advantages to the deferral of wage income. One advantage derives from giving labor a stake in the profitability of the firm. Ippolito (1985) has developed a model in which pension contracts reduce the incentive of labor to exploit its monopoly in specialized skills in the short run, at the expense of the long-run survival of the firm. By making employees creditors of the firm, a congruence of goals between capitalists and labor is established. In a separate vein, firms that sponsor defined-benefit pension plans also realize tax advantages. Employer contributions to the pension fund are tax-deductible and are exempt from payroll taxes. Finally, borrowing from employees may be a more flexible source of financing than selling bonds.

The model incorporates the heterogeneous nature of the federal personal income tax. Let $t(pd, i)$ and $t(pe, i)$ represent the constant marginal tax rates on debt and equity income for individual i . The opportunity to defer the recognition of capital

gains indefinitely causes the tax rate on debt income to exceed the effective rate on equity income, implying that $t(pd, i) > t(pe, i)$. That relationship creates a personal tax bias against debt income which will ultimately influence the equilibrium pension benefit.

The corporate tax rate is a constant $t(c)$. Marginal tax rates on investment income are assumed to be progressive in nature, with at least one investor in each of the following mutually exclusive and exhaustive tax brackets:

Bracket 1 $(1 - t(pd, i)) > (1 - t(c))(1 - t(pe, i))$

Bracket 2 $(1 - t(pd, i)) = (1 - t(c))(1 - t(pe, i))$

Bracket 3 $(1 - t(pd, i)) < (1 - t(c))(1 - t(pe, i))$

At the end of the one-period time frame, corporations are assumed to receive state dependent operating earnings of $X(s)$ in state s ; $X(s)$ is monotonically increasing in s with $0 < X(S1) < X(S5)$. The face value of debt (B), certain noncash charges (D) such as depreciation and depletion, and pension contributions (VP) are excludable from taxable income. Firm specific tax credits of G dollars exist, but consistent with the tax code only $\theta\%$ of a firm's tax liability can be offset by these credits.

87.7.1 Aggregate Pension Demand and Differential Taxation

This section explains the characteristics of the demand for pension income in the model. That demand is determined in the market for the nonrecourse debt securities that are collateralized by pension income. The model mirrors the fundamental structure of the market for corporate debt and equity claims developed by Miller (1977) and DeAngelo and Masulis (1980).

To begin, recall that $t(pd, i)$ and $t(pe, i)$ represent the constant marginal personal tax rates on pension collateralized debt and equity income, respectively, for the i th individual. The opportunity to defer the recognition of capital gains introduces a bias into the tax code, so that $t(pd, i) > t(pe, i) > 0$ for all i . That implies that the state-contingent personal after-tax cash flow per unit of state s equity income exceeds that per unit of state s debt income, so that $(1 - t(pe, i)) > (1 - t(pd, i))$ for all individuals.

Let $P(d, s)$ equal the current time ($t = 0$) price in the market of \$1 of pretax state s income from an investment in pension collateralized debt. The term $P(e, s)$ represents the current time market price of \$1 of pretax state s income from an equity investment. Their reciprocals, $1/P(d, s)$ and $1/P(e, s)$, are pretax current yields; $(1 - t(pd, i))/P(d, s)$ and $(1 - t(pe, i))/P(e, s)$ are accordingly the after-tax yields to the i th individual. Utility maximizing investors will adjust their portfolio holdings of states claims in order to maximize these after-tax yields. In making their portfolio selection, investors will prefer pension collateralized debt to equity if $(1 - t(pd, i))/P(d, s) > (1 - t(pe, i))/P(e, s)$; equity will be preferred to debt if the inequality is reversed; finally, the individual will be indifferent between the two when after-tax yields are equal.

Individual members of the labor pool function as intermediaries in the pension contracting process. Let $g(t(w, i))$ represent the value of deferring \$1 of income from time 0 to time 1, for the i th individual with a time 0 marginal wage tax rate of $t(w, i)$. The schedule of marginal tax rates on wage earnings ($w(i)$) is progressive, such that $t(w, i)$ increases with $w(i)$. Define r to be the opportunity cost of funds. The value of deferring the tax on \$1 of wage income is $g(t(w, i)) = t(w, i) * (r/(1 + r))$. The function $g(t(w, i))$ is nonnegative, and is monotonically increasing in $t(w, i)$.

The intermediation process operates in the following way. Individuals are willing to exchange $\$1/(1 + r)$ of time 0 wage income for \$1 of pension income at the end of the period. As members of the labor group, however, they desire to neutralize the financial holding (i.e., pension obligation) created by that exchange. To accomplish this, each individual must pay a certification premium of $\$c$ to establish credentials in the capital market. Once certified, any individual may sell a personal debt claim (collateralized by the pension benefit) on a nonrecourse basis. The sale of debt by the employee offsets the holding of the pension claim, leaving him/her with a (desired) zero net investment position.

The cross-sectionally constant corporate tax rate is represented by $t(c)$. It is assumed that progressive personal tax rates are such that at least one investor is in each of the following mutually exclusive and exhaustive tax brackets:

Debt Preferring $(1 - t(pd, i)) > (1 - t(c))(1 - t(pe, i))$

Debt Indifferent $(1 - t(pd, i)) = (1 - t(c))(1 - t(pe, i))$

Debt Adverse $(1 - t(pd, i)) < (1 - t(c))(1 - t(pe, i))$

Investors are risk neutral and homogeneous in their belief that each state will occur with probability $p(s)$. In the course of constructing a portfolio, every investor will select the security of his/her tax preference (pension collateralized debt or equity) with the highest expected yield. Given that all claims must be held in equilibrium, it follows that the markets must set prices for their state contingent claims so that pretax expected yields are equal. If $P(d)$ and $P(e)$ are the current market prices of the pretax cash flows from debt and equity, respectively, this implies that $1/P(d) = p(s)/P(d, s) = p(s)/P(e, s) = 1/P(e)$ for all s . Applying the results of DeAngelo and Masulis (1980), it is possible to examine both (1) the pension collateralized debt and equity demand decisions of individuals, and (2) the pension and equity supply decisions of firms by examining only two markets: the one for pretax cash flows to equity (with current price $P(e)$) and the one for pretax cash flows to pension collateralized debt (with current price $P(d)$). Let the superscript u denote investors who are indifferent between investing in debt or equity. Exactly which investors are marginal in this sense is established by the market prices $P(d)$ and $P(e)$, when they equate the pretax personal yields of pension collateralized debt and equity for certain groups of investors. This occurs when, for individual u :

$$p(s)(1 - t(pd, u))/P(d, s) = (1 - t(pd, u))/P(d) = (1 - t(pe, u))/P(e) = p(s)(1 - t(pe, u))/P(e, s).$$

Equivalence of personal after-tax yield causes individuals to be indifferent between investing their next dollar of capital in pension collateralized debt or equity.

87.7.2 Firm Valuation

This section describes the one-period state preference valuation model that will be used in the subsequent discussion of optimal corporate pension policy. The model is an adaptation of the theoretical work of DeAngelo and Masulis (1980). It is from this valuation model that the supply function for pension income emerges.

Let $X(s)$ represent cash operating earnings before pension expenditures. This variable is monotonically increasing in s , with $X(S1) < X(s) < X(S5) < \infty$. From $X(s)$, certain noncash charges (D), the interest expense on bonds (B), and vested pension benefits (VP) are deducted to arrive at corporate taxable income. Firms pay a tax of $t(c)$ on taxable earnings. They can offset their liability with any tax credits (G) which they possess, subject to the limitation that the offset does not exceed $\theta\%$ of the total liability. The firm-specific variables B , VP , D , and G define the subintervals over the $[S1, S5]$ range that describe the potential pretax dollar returns which the holders of pension collateralized debt and equity face. Table 1 describes those returns, which will be referred to as $VP(s)$ and $E(s)$, respectively.

In the first interval, $[S1, S2]$, the operating stream is sufficient to compensate regular bondholders but is insufficient or just sufficient to pay pension benefits. The pension beneficiaries will receive from zero to VP dollars, while the holders of the nonrecourse debt sold on those claims will receive the after-tax proceeds to the beneficiaries. The firm will pay no taxes, thereby wasting all deductions and credits.

When $(X(s) - B) < X(s) < (B + VP + D)$, then $s \in [S2, S3]$. In this interval, both bondholders and retirees are paid in full. No corporate taxes are paid, however, as earnings fail to exceed total tax deductions. Excess deductions and all tax credits are consequently unutilized. When $(VP + B + D) < X(s) < (VP + B + D) + G/\theta t(c)$, $s \in [S3, S4]$. This interval is similar to the prior one, except that the firm now has earnings that exceed all available deductions, resulting in a tax liability that can be partially offset by available tax credits. The limitation on the offset exists because the total credits available exceed the maximum percentage of the gross liability allowed by law. Thus, all deductions and a fraction of available credits are utilized. In the final interval, when $s \in [S4, S5]$, the returns to both parties reflect the full utilization of all tax deductions and tax credits by the firm.

The optimal deferred wage occurs when the current time firm value is maximized, that is when the market value (V) of the claims of labor (VP) and equity (E) are greatest in total. VP and E are valued by discounting their cash flows in every state nature by the appropriate state-contingent unit price, $P(d, s)$ and $P(e, s)$, respectively. Thus,

Table 1 State Contingent Dollar Returns

State	$VP(s)$	$E(s)$
$S1-S2$	$X(s) - B$	0
$S2-S3$	VP	$X(s) - B - BP$
$S3-S4$	VP	$S(s) - B - VP - (1 - \theta)t(c)[X(s) - B - VP - D]$
$S4-S5$	VP	$S(s) - B - VP - t(c)[X(s) - B - VP - D] + G$

$$\begin{aligned}
V &= VP + EV = \int_{S1}^{S2} VP(s)P(d, s)ds + \int_{S1}^{S5} E(s)P(e, s)ds \\
&= \int_{S1}^{S2} [X(s) - B]P(d, s)ds + \int_{S2}^{S5} [VP]P(d, s)ds \\
&\quad + \int_{S2}^{S3} [X(s) - B - VP]P(e, s)ds \\
&\quad + \int_{S3}^{S4} [(X(s) - B - VP) - (1 - \theta)t(c)(X(s) - VP - B - D)]P(e, s)ds \\
&\quad + \int_{S4}^{S5} [X(s) - B - VP - t(c)((X(s) - B - VP - D) + G)]P(e, s)ds \quad (1)
\end{aligned}$$

The objective of the firm is to select the VP^* value that maximizes V . The appropriate first-order condition is:

$$\begin{aligned}
\partial V / \partial VP &= \int_{S2}^{S3} [P(d, s) - P(e, s)]ds \\
&\quad + \int_{S3}^{S4} [P(d, s) - P(e, s)(1 - (1 - \theta)t(c))]ds \\
&\quad + \int_{S4}^{S5} [P(d, s) - P(e, s)(1 - t(c))]ds \quad (2)
\end{aligned}$$

Equation (2) shows that the level of unfunded pension liabilities which a firm possesses is relevant to valuation because $\partial V / \partial VP$ is not equal to zero for all pension policies if D , B , and/or G are positive. That implies that some values of VP are strictly preferred to others.

Under the assumption that investors are risk-neutral and have homogeneous beliefs, Eq. (2) reduces to

$$\begin{aligned}
\partial V / \partial VP &= [P(d) - P(e)] \int_{S2}^{S3} P(s)ds \\
&\quad + [P(d) - (1 - (1 - \theta)t(c))P(e)] \int_{S3}^{S4} P(s)ds \\
&\quad + [P(d) - (1 - t(c))P(e)] \int_{S4}^{S5} P(s)ds \quad (3)
\end{aligned}$$

In the aggregate, Eq. (3) represents the present value of the expected corporate after-tax cash flow that results at the margin when unfunded pension liabilities are substituted for equity in the firm's financial structure. Further, this marginal present value contains three distinct components whose magnitudes depend on the degree to

which the marginal pension deduction is utilized. The first term, $P(d) - P(e)(1 - t(c))$, is the present value of the unfunded pension for equity substitution when the marginal corporate tax deduction resulting from the increased pension liability is fully utilized. The corresponding integral represents the probability of full utilization of the marginal pension deduction. The second term $P(d) - P(e)(1 - (1 - \theta)t(c))$, is a lower present value due to the partial loss of the corporate tax shield caused by the statutory 0 ceiling on usable tax credits. The corresponding integral is the probability of the partial loss of the marginal pension deduction due to the 0 ceiling. Finally, the third term, $P(d) - P(e)$, represents the present value of the unfunded liability substitution when available deductions already exceed total earnings. The third integral is the probability of the total loss of the deduction for the marginal pension.

87.7.3 Equilibrium with No Tax Shield Risk

To show the contribution of this model to the body of pension theory, this section will adapt the DeAngelo and Masulis market equilibrium in a way that allows for the derivation of analytic results similar to those obtained by Sharpe (1976). When there are no tax shields other than the pension contribution, the partial or total loss of the marginal tax benefits of sponsoring a plan is impossible. If $D = B = G = 0$, then $S2 = S3 = S4$, and Eq. (3) reduces to

$$\partial V / \partial VP = [P(d) - (1 - t(c))P(e)] \int_{S2}^{S5} P(s) ds \quad (4)$$

Equation (4) allows for the derivation of the aggregate supply curve for pension income.

Turning briefly to the demand side, marginal investors must be in bracket 2 at market equilibrium. As pointed out in Sect. 87.7.1, marginal investors are those for whom personal after-tax expected yields on debt and equity are equal: $(1 - t(pd, u))/P(d) = (1 - t(pe, u))/P(e)$. When $P(d) = P(e)(1 - t(c))$, marginal investors are in bracket 2 with tax rates satisfying $(1 - t(pd, i))(1 - t(pe, i))(1 - t(c, i))$.

Substituting the marginal investor's tax rate condition $P(d) = P(e)(1 - t(pe, u))/(1 - t(pd, u))$ into the marginal value of debt Eq. (4) yields irrelevant to firm value.

$$\begin{aligned} \partial V / \partial VP[VP^*] &= \left[\frac{P(d)}{1 - t(pd, u)} \right] * [1 - t(pd, u)) - (1 - t(pe, u)) \\ &\quad \times (1 - t(c))] \int_{S2}^{S5} p(s) ds \end{aligned} \quad (5)$$

In equilibrium, the expression in brackets is zero because marginal investors are in bracket 2. The market in an intuitive sense endogenously determines the relative

marginal personal tax rates on debt and equity so that the corporate pension advantage exactly offsets the disadvantage of pension collateralized debt to investors at any debt level.

The model, therefore, shows that pension funding policy will be irrelevant to firm value when pension contributions are always deductible. This is because, at the margin, the constant corporate tax advantage of providing more pension income is offset by the personal tax disadvantage of doing so. While an optimal level of unfunded pensions will exist in the economy, there are no individual optimums which obtain that are unique to each firm. In this instance, the equilibrium characteristics are the same as those found in Sharpe (1976). The next section will explain how the existence of non-pension tax shields affects the analysis.

87.7.4 Tax Shield Risk and Optimal Corporate Pension Policy

When corporate tax shield substitutes exist ($D, B, G > 0$), corporate pension policy is no longer irrelevant. Instead, relative market prices will adjust until each firm has a unique interior optimum pension level, in equilibrium. Similar to the previous section, there is a constant expected marginal personal tax disadvantage to debt. The existence of tax shield substitutes, however, causes the expected marginal corporate tax benefit to decline as more pension income is contracted. At VP^* , the expected marginal corporate tax benefit just equals the expected marginal personal tax disadvantage of the pension collateralized debt. Similar to the findings of DeAngelo and Masniis, unique interior optimums exist when firms have different amounts of debt substitutes.

To begin, the aggregate supply curve for pension income will be derived, under the assumption that the pension income stream is riskless. When pension contributions are not the firm's only tax shield, both individual and aggregate supply curves will consist of a perfectly elastic section, after which they will smoothly slope upward. The curve is perfectly elastic at relative prices $P(d) = P(e)(1 - t(c))$ over the levels of pension income which allow for the full utilization of all corporate tax shields. The supply curve is upward sloping beyond the full utilization level because firms are willing to supply more pension income to labor only if they are compensated (for the greater likelihood of a partial or total loss of the additional tax shield) with higher unit pension collateralized debt prices $P(d) > P(e)(1 - t(c))$. The supply curve represents the schedule of $(P(d), VP)$ combinations at which the firm has unique interior optimum pension levels. In market equilibrium, the point at which the aggregate demand curve intersects the aggregate supply curve (in its upward sloping section, under weak assumptions on the personal tax code) sets relative prices $P(d) > P(e)(1 - t(c))$. These prices in turn determine the unique interior optimum pension policy for each firm.

A formal derivation of the supply curve proceeds as follows. When D, B , and $G > 0$, $S_4 > S_3 > S_2 > 0$, and the first-order condition is once again represented by Eq. (3). If relative prices are such that $P(d) < P(e)(1 - t(c))$, $\partial V / \partial VP < 0$ for

all values of VP , and the firm will supply no pension income. When $P(d) = P(e)(1 - t(c))$, the supply curve is perfectly elastic to the point where the selected pension contribution just allows for the full utilization of all tax shields in every state of nature. Referring to this amount as $PMAX$, $PMAX$ occurs when θ times the gross tax liability is greater than or equal to G . This is written in equation form as

$$\theta t(c)[X(S1) - D - B - PMAX] > G$$

which, when solving for $PMAX$ implies

$$PMAX = X(S1) - D - B - \frac{G}{\theta t(c)} < X(S1)$$

When VP is such that $0 < VP < PMAX$, $S2 = S3 = S4$ because all corporate tax shields are fully utilized in every state of nature. The first-order condition reduces to

$$\partial V / \partial VP = P(d) - P(e)(1 - t(c)) \text{ for } 0 < VP < PMAX$$

At relative prices $P(d) = P(e)(1 - t(c))$, the firm is indifferent to all pension levels that allow for the full utilization of all corporate tax shields. Consequently, the supply curve is perfectly elastic at relative prices $P(d) = P(e)(1 - t(c))$ over the interval $0 < VP < PMAX$.

No pension income will be provided beyond $PMAX$ at relative prices $P(d) = P(e)(1 - t(c))$. That is because when $PMAX < VP < X(S1)$, $S4 > S3 > S2$. If $P(d) = P(e)(1 - t(c))$, $\partial V / \partial VP < 0$, because the second and third terms are negative, while the first is zero. Firms will only provide pension income beyond $PMAX$ if $\partial V / \partial VP > 0$. In the context of Eq. (3), that is possible only if $P(d) > P(e)(1 - t(c))$. From an economic perspective, the higher pension price (rendered in the form of greater wages returned) is necessary in order to compensate shareholders for the loss of the corporate tax shield which they will incur on marginal units of pension contribution when $s \in [S3, S4]$. For all prices $P(d) > P(e)(1 - t(c))$, there exists a VP^* which sets $\partial V / \partial VP = 0$. As $P(d)$ rises relative to $P(e)$, Eq. (3) will become positive, and the firm will find it advantageous to supply more pension income. The supply curve is thus smoothly upward

sloping beyond $PMAX$.

As DeAngelo and Masulis (1980) point out, market equilibrium will occur on the upward sloping portion of the pension supply curve under a set of weak assumptions. First, the perfectly elastic section of the supply curve must be short. This is reasonable because few pension levels are both riskless to the employee and which have a zero probability of tax shield loss, as the range $0 < VP < PMAX$ does. Second, investors in bracket 1 must be sufficient in number to demand larger quantities of debt than $PMAX$ in the aggregate, at relative prices $P(d) = P(e)(1 - t(c))$. If both of these conditions hold, then the aggregate demand curve will intersect the aggregate supply curve at relative prices $P(d) > P(e)(1 - t(c))$. These prices will dictate a unique

interior optimum pension policy for each firm, and imply that investors are in bracket 1 with tax rates satisfying $(1 - t(pd, i)) > (1 - t(pe, i))(1 - t(c))$.

Following DeAngelo and Masulis, the first-order condition can be used to intuitively show the trade-off that determines the optimal pension policy. For the sake of simplifying the analytics, it is assumed that $t(pe, i) = t(pe, u) = 0$, and that VP^* occurs over the range $PMAX < VP^* < X(S1)$. Evaluating the first-order condition at VP^* and substituting the simplified marginal personal tax rate condition $P(e) = P(d)/(1 - t(pd, u))$ yields

$$\frac{\partial V}{\partial VP[VP^*]} = \left[\frac{P(d)}{1 - t(pd, u)} \right] * \left[t(c) \int_{S4}^{S5} p(s) ds + (1 - \theta) \int_{S3}^{S4} p(s) ds \right] - t(pd, u) \quad (6)$$

The first term in square brackets represents the expected corporate tax advantage of an additional dollar of pension income substituted for equity. Notice that this term is higher when the probability of losing any or all of the marginal tax shield is lower. The second term in square brackets, $t(pd, u)$, quantifies the expected marginal personal tax disadvantage of riskless debt. This is because, under the simplifying assumptions made, an increase in VP of one unit raises the personal tax liability by $t(pd, u)$ in all states of nature. At low levels of pension income, the first term exceeds the second term because there is a higher probability that the additional pension contribution can be fully utilized to reduce the firm's tax liability, and investors purchasing the corresponding debt are in the lower tax brackets. As the quantity of pension income provided rises, so does the likelihood that any or all of the marginal tax benefit will be lost. At equilibrium, the declining marginal corporate advantage of pension contributions just equals the marginal personal disadvantage of the debt sold on pension income. At VP^* , it is not advantageous for firms to supply more pension income or for labor to demand more.

As is the case in the DeAngelo and Masulis model, the use of a one-period framework necessarily precludes the consideration of the carry-back and carry-forward (CB-CF) provisions in the tax code. The effect of the inclusion in a multi-period model is not difficult to predict. Both reduce the probability that a corporate tax shield will be lost. CB provisions allow any unused shields to be applied to reduce the tax liability of up to three prior years. CF provisions allow the firm to apply any shields which cannot be carried back to up to 15 years in the future. Because both involve a time-value loss, however, they reduce but do not eliminate, the expected value of the corporate tax shield loss on the marginal unit of pension. Firms would therefore require less price compensation for increasing the supply of pension income. As a result, the supply curve in a multi-period model would be more elastic in its upward sloping portion than the one presented here.

Finally, it should be noted that our analysis of pension policy can be generalized to the case when corporate bonds are risky. Under that assumption, the returns to the contributors of capital are described in Table 2, given that bondholder claims outrank those of unfunded pensions in a liquidation. Following Table 2, if $P(b, s)$ represents

Table 2 State Contingent Dollar Returns

<i>State</i>	<i>B(s)</i>	<i>VP(s)</i>	<i>E(s)</i>
<i>S1-S2</i>	<i>X(s)</i>	0	0
<i>S2-S3</i>	<i>B</i>	<i>X(s) - B</i>	0
<i>S3-S4</i>	<i>B</i>	<i>VP</i>	<i>X(s) - B - VP</i>
<i>S4-S5</i>	<i>B</i>	<i>VP</i>	<i>X(s) - B - VP - (1 - θ)t(c)[X(s) - B - VP - D]</i>
<i>S5-S6</i>	<i>B</i>	<i>VP</i>	<i>X(s) - B - VP - t(c)[X(s) - B - VP - D] + G</i>

the price per unit of state-contingent income to corporate bonds, then the valuation expression can be written as

$$\begin{aligned}
 V = & \int_{S1}^{S2} [X(s)P(b, s)]ds + \int_{S2}^{S4} [B]P(b, s)ds + \int_{S2}^{S3} [X(s) - B]P(d, s)ds \\
 & + \int_{S3}^{S6} [VP]P(d, s)ds + \int_{S3}^{S4} [X(s) - B - VP]P(e, s)ds \\
 & + \int_{S4}^{S5} [X(s) - B - VP - (1 - \theta)t(c)(X(s) - B - VP - D)]P(e, s)ds \\
 & + \int_{S5}^{S6} [X(s) - B - P - t(c)(X(s) - B - P - D) + G]P(e, s)ds
 \end{aligned}$$

and the appropriate first-order condition is

$$\begin{aligned}
 \partial V / \partial VP = & \int_{S5}^{S6} [P(d, s) - P(e, s)(1 - t(c))]ds \\
 & + \int_{S4}^{S5} [P(d, s) - P(e, s)(1 - (1 - \theta)t(c))P(e, s)]ds \\
 & + \int_{S3}^{S4} [P(d, s) - P(e, s)]ds
 \end{aligned}$$

Under the special conditions of risk neutrality and homogenous beliefs,

$$\begin{aligned}
 \partial V / \partial VP = & [P(d) - P(e)] \int_{S3}^{S4} p(s)ds + [P(d) - P(e)(1 - (1 - \theta)t(c))] \int_{S4}^{S5} p(s)ds \\
 & + \left[P(d) - P(e)(1 - t(c)) \right] \int_{S5}^{S6} p(s)ds
 \end{aligned}$$

The first-order condition is of the same form as in the case of riskless debt but differs with regard to the probabilities of the respective state intervals. That is

because $\int_{s1}^{s2} p(s) ds > 0$ in the case of risky conventional debt, implying that the supply curve has a greater slope than when conventional debt is riskless.

87.8 Conclusion

This study examines how pension accounting relates to inside debt and capital structure. First, as defined benefit (DB) and defined contribution (DC) plans are the two most common types of pension plans, we demonstrate that DC plans are cheaper and involve lower risk for sponsoring firms than DB plans. Therefore, the shift in pension coverage from DB to DC plans has gained popularity among large healthy firms since the early 2000s (Munnell et al. 2007). Second, by using an illustration with Fig. 1, we explain how Employee Retirement Income Security Act (ERISA) of 1974 improves the protection of retirement benefits for private-sector employees and offers a put option for sponsoring firms. Third, we provide a review of empirical research on the pension accounting associated with earnings management and firm value. We highlight that institutional settings difference leaves rooms for future empirical research on pension accounting. Finally, following Alderson and Lee (1988), this chapter introduces a developed model in which individuals and corporations have an incentive to engage in contracts for the deferral of wage income in the form of an unfunded pension as part of the capital structure decision.

References

- Alderson, M.J., and C.F. Lee. 1988. Corporate pension policy and capital structure decisions. *Journal of Economics and Business* 40: 209–228.
- Amir, E., and E.A. Gordon. 1996. Firms' choice of estimation parameters: Empirical evidence from SFAS 106. *Journal of Accounting, Auditing and Finance* 3: 427–448.
- Amott, R., and A. Gersovitz. 1980. Corporate financial structure and the functioning of private pension plans. *Journal of Public Economics* 2: 231–248.
- Barth, M., W.H. Beaver, and W.R. Landsman. 1993. A structural analysis of pension disclosures under SFAS 87 and their relation to share prices. *Financial Analysts Journal* 49: 18–26.
- Bergstresser, D.B., M.A. Mihir, and J. Rauh. 2006. Earnings manipulation, pension assumptions and managerial investment decisions. *The Quarterly Journal of Economics* 121: 157–195.
- Borah, N.B., H.L. James, and J.C. Park. 2020. Does CEO inside debt compensation benefit both shareholders and debtholders? *Review of Quantitative Finance and Accounting* 54: 159–203.
- Calhoun, C.V. 2010. Maximum benefits and Contributions Limits for 2005–2010. Available at <http://benefitsattorney.com/modules.php?name=415>.
- Cassell, C.A., S.X. Huang, J.M. Sanchez, and M.D. Stuart. 2012. Seeking safety: The relation between CEO inside debt holdings and the riskiness of firm investment and financial policies. *Journal of Financial Economics* 103: 588–610.
- Coronado, L.Y., and S.A. Sharpe. 2003. Did pension plan accounting contribute to a stock market bubble? *Brookings Papers on Economic Activity* 1: 323–371.
- Coronado, J., O.S. Mitchell, S.A. Sharpe, and S.B. Nesbitt. 2008. Footnotes aren't enough: the impact of pension accounting on stock values. *Journal of Pension Economics and Finance* 7: 257–276.

- Choy, H., J. Lin, and M.S. Officer. 2014. Does freezing a defined benefit pension plan asset firm risk? *Journal of Accounting and Economics* 57: 1–21.
- Colonnello, S., G. Curatola, and N.G. Hoang. 2017. Direct and indirect risk-taking incentives of inside debt. *Journal of Corporate Finance* 45: 428–466.
- Comprix, J., and K.A. Muller. 2006. Asymmetric treatment of reported pension expense and income amounts in CEO cash compensation calculations. *Journal of Accounting and Economics* 42: 385–416.
- . 2011. Pension plan accounting estimates and the freezing of defined benefit pension plans. *Journal of Accounting and Economics* 51: 115–133.
- Chi, S., S.X. Huang, and J.M. Sanchez. 2017. CEO inside debt incentives and corporate tax sheltering. *Journal of Accounting Research* 55: 837–876.
- Dale, A. 2003. Pension fund losses to be revealed. *The Salt Lake Tribune*, January 16.
- Davis-Friday, P.Y., L.B. Folami, C. Liu, and H.F. Mittelstaedt. 1999. The value relevance of financial statement recognition versus Disclosure: evidence from SFAS No. 106. *The Accounting Review* 74: 403–423.
- DeAngelo, H., and R. Masulis. 1980. An optimal capital structure under corporate and personal taxation. *Journal of Financial Economics* 8: 3–30.
- Franzoni, F.A., and J.M. Martin. 2006. Pension plan funding and stock market efficiency. *Journal of Finance* 61: 921–956.
- Freund, S., S. Latif, and H.V. Phan. 2018. Executive compensation and corporate financing policies: Evidence from CEO inside debt. *Journal of Corporate Finance* 50: 484–504.
- Glaum, M. 2009. Pension accounting and research: A review. *Accounting and Business Research* 39: 273–311.
- Godwin, J.H., S.R. Goldberg, and J.E. Duchac. 1996. An empirical analysis of factors associated with changes in pension-plan interest-rate assumptions. *Journal of Accounting, Auditing and Finance* 11: 305–322.
- Gopalakrishnan, V., and T.F. Sugrue. 1995. The determinants of actuarial assumptions under pension accounting disclosures. *Journal of Financial and Strategic Decisions* 8: 35–41.
- Hann, R.N., Y.Y. Lu, and K.R. Subramanyam. 2007a. Uniformity versus flexibility: Evidence from pricing of the pension obligation. *The Accounting Review* 82: 107–137.
- Hann, R., F. Hefflin, and K.R. Subarmanayam. 2007b. Fair-value pension accounting. *Journal of Accounting and Economics* 44: 328–358.
- Hsieh, S.J., Y. Su, and C.A. Chang. 2021. The role of discretionary pension accruals in earnings management. *Journal of Applied Accounting Research*. forthcoming.
- Hsu, A.W., C.F. Wu, and J.C. Lin. 2013. Factors in managing actuarial assumptions for pension fair value: Implications for IAS 19. *Review of Pacific Basin Financial Markets and Policy* 16: 1350002.
- Ippolito, R. 1985. The economic function of underfunded pension plans. *Journal of Law and Economics* 28: 611–651.
- Kieso, D.E., J.J. Weygandt, and T.D. Warfield. 2020. *Intermediate accounting IFRS*. 4th ed. Wiley.
- Landsman, W.R. 1986. An empirical investigation of pension fund property right. *The Accounting Review* 61: 622–691.
- Landsman, W., and J. Ohlson. 1990. Evaluation of market efficiency for supplementary accounting disclosures: The case of pension assets and liabilities. *Contemporary Accounting Research* 7: 185–198.
- Lee, C.F., C. Hu, and M. Foley. 2021. Differential risk effect of inside debt, CEO compensation diversification, and firm investment. *Review of Quantitative Finance and Accounting* 56: 506–543.
- Lin, J.J., and E.M. Werner. 2021. Discretionary pension assumption choices, corporate governance, and earnings management: evidence from Taiwan. *Advances in Quantitative Analysis of Finance and Accounting* 18: 139–169.
- McFarland, B., G. Pang, and M. Warshawsky. 2009. Does freezing a defined-benefit pension plan increase company value? Empirical evidence. *Financial Analysts Journal* 65 (4): 47–59.

- Miller, M.H. 1977. Debt and taxes. *Journal of Finance* 32: 261–276.
- Munnell, A.H., F. Golub-Sass, M. Soto, and F. Vitagliano. 2007. Why are healthy employers freezing their pensions? Boston College Center for Retirement Research, 44.
- Picconi, M. 2006. The perils of pensions: does pension accounting lead investors and analysts astray? *The Accounting Review* 81: 925–955.
- Sundaram, R.K., and D.L. Yermack. 2007. Pay me later: Inside debt and its role in managerial compensation. *Journal of Finance* 62: 1551–1588.
- Sharpe, W. 1976. Corporate pension funding policy. *Journal of Financial Economics* 4: 183–193.
- Wang, F.A., and T. Zhang. 2014. The effect of unfunded pension liabilities on corporate bond ratings, default risk, and recovery rate. *Review of Quantitative Finance and Accounting* 43 (4): 781–802.
- Watson W. 2010. Pension freezes continue among Fortune 1000 companies in 2010. *Insider* (September): 1–4.
- Wei, C., and D. Yermack. 2011. Investor reactions to CEOs' inside debt incentives. *Review of Financial Studies* 24: 3813–3840.
- Wiedman, C.I., and H.A. Wier. 2004. The market value implications of post-retirement benefit plans and plan surpluses – Canadian evidence. *Canadian Journal of Administrative Sciences* 21: 229–241.



The Role of Earnings Management in Equity Valuation **88**

Guanming He, April Zhichao Li, and Dongxiao Shen

Abstract

The intrinsic value of a firm's equity is determined by the present value of future payoffs to equity-holders. Thus, to estimate equity value, one needs to identify and process a series of information that is relevant to the present value of expected future payoffs. Among the value-relevant information, earnings pertain to a summary measure of firm performance and thus play a key role in the equity valuation. Nevertheless, it is not uncommon in practice that firm managers either intentionally or unintentionally misreport earnings, making valuation challenging for outsiders. This chapter aims to explore the role earnings management plays in equity valuation, and how market participants may detect and adjust for misreported earnings, if any, in their forecasts and valuation for a firm. The chapter is organized to comprise the following sections: (i) varied managerial incentives for earnings management; (ii) various tactics of earnings management via accruals manipulation; (iii) various tactics of earnings management via real activities manipulation; (iv) consequences and determinants of earnings management; (v) trade-off between accruals manipulation and real activities manipulation to manage earnings; (vi) how to discern and measure earnings management; (vii) introduction of various equity-valuation models; and (viii) how to adjust for the effects of accruals-based and real earnings management in equity valuation.

Keywords

Accruals-based earnings management · Real earnings management · Equity valuation

G. He (✉) · A. Z. Li · D. Shen

Department of Accounting, Durham University, Durham, UK

e-mail: guanming.he@durham.ac.uk; zhichao.li@durham.ac.uk; dongxiao.shen@durham.ac.uk

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_90

2061

JEL Classification

M41 · G10 · G17

88.1 Varied Managerial Incentives for Earnings Management

Investors tend to focus on earnings to make their investment decisions (e.g., Biddle et al. 1995; Francis et al. 2003; Graham et al. 2005). Therefore, managers can manipulate earnings to mislead investors' perception about the firm's future prospects. There are various scenarios in which managerial incentives for earnings management are aligned with shareholders' interests. We classify these scenarios by upwards *vis-à-vis* downwards earnings management, and expatiate on them below.

To facilitate external financing, managers manage earnings upwards prior to initial public offerings or seasoned equity offerings (Rangan 1998; Teoh et al. 1998; Shivakumar 2000; Haw et al. 2004). In so doing, managers can boost the market expectation of a firm's future performance, thereby selling shares at a higher price and raising more capital at a lower cost.

Prior to the announcements of stock-for-stock mergers, managers in both acquirers and acquirees are inclined to manipulate earnings upwards to increase acquisition premium which is tied to stock price (Erickson and Wang 1999; Louis 2004). The higher the stock prices of acquirers before the stock mergers, the fewer stocks acquirers need to swap with acquirees, and vice versa.

Most debt covenants that firms are required to meet are contingent on financial ratios (e.g., interest coverage ratio) that involve earnings numbers. Violation of the covenants can trigger penalties, including acceleration of debt repayment and increases in interest rates, for the firms. Thus, when firms are close to violating debt covenants, managers have an incentive to manage earnings to avoid covenant violation (DeFond and Jiambalvo 1994; Sweeney 1994; Beneish et al. 2012).

Firms missing the earnings thresholds such as zero earnings or consensus analyst earnings forecasts would likely have a stock-price plunge. On the contrary, firms consistently meeting or beating the earnings thresholds would likely receive a price premium (e.g., Barth et al. 1999; Kasznik and McNichols 2002; Myers et al. 2007; Bhojraj et al. 2009). Therefore, managers have a propensity to manipulate earnings upwards to meet or beat the earnings thresholds (Burgstahler and Dichev 1997; Robb 1998; Schilit 2010).

Given the limited amount of internal funds available for investments or operations, financially constrained or financially distressed firms need more external funds (He and Ren 2021). To facilitate external financing for relieving financial constraints or distress, managers have a tendency to manage earnings upwards (Loebbecke et al. 1989; Maksimovic and Titman 1991). Firms in financial constraints or distress are often poorly performing and thus have less to lose from getting caught with financial misconduct (Amiram et al. 2018); this further strengthens managers of these firms to inflate earnings.

Before firms repurchase stocks from open stock markets, managers of these firms have an incentive to deflate earnings to suppress the price paid for the repurchase (e.g., DeAngelo 1986; Gong et al. 2008). The deflated repurchase price transfers wealth from the shareholders who sell (i.e., the leaving shareholders) to those who hold onto their shares (i.e., the remaining shareholders). The lower the price firms manage to suppress through earnings deflation, the more benefits the remaining shareholders would obtain through this wealth transfer; so too is for the executives who hold a substantial amount of stocks or stock options. The likelihood of obtaining import relief and the amount of relief granted by governments will increase if a firm can show that its profitability was adversely tainted by foreign competition. As such, firms seeking import relief have a tendency to make profit-decreasing accounting discretion (Jones 1991). Firms are also likely to make profit-decreasing accounting choices to tone down their performance to outsiders when facing fierce industrial product market competitions or high threats of new entrants (Cheng et al. 2013; Balakrishnan and Cohen 2013; Dhaliwal et al. 2014).

Earnings management could be merely in the managers' own interests and not be aligned with the interests of incumbent shareholders. In such cases, managers' incentives for earnings management are notably stronger. For instance, the value of stocks (stock options) held by managers is contingent on the level (volatility) of stock prices (Coles et al. 2006; Low 2009; Armstrong et al. 2013). Thus, managers holding stocks and stock options have an incentive to manipulate earnings to maximize the value of their equity-based compensation (Cheng and Warfield 2005; Bergstresser and Philippon 2006; Burns and Kedia 2006; Erickson et al. 2006; Efendi et al. 2007; Cheng and Farber 2008; Cornett et al. 2008; Armstrong et al. 2010; Jiang et al. 2010; Feng et al. 2011; Armstrong et al. 2013). The amount of bonus pay is often tied to reported earnings in excess of a pre-specified earnings level. The higher the earnings reported by managers, the higher bonus pay they can get from their companies. This incentivizes the managers to manage earnings to enlarge bonus pay (Healy 1985; Holthausen et al. 1995).

Given high stock holdings by managers, they have an incentive to manage earnings upwards to inflate stock prices before selling shares they own (Ke et al. 2003; Beneish 1999; Park and Park 2004; Darrough and Rangan 2005; Beneish and Vargus 2002; Sawicki and Shrestha 2014). In the same vein, managers are prone to manipulate earnings downwards to deflate stock prices before purchasing shares from open stock markets (Sawicki and Shrestha 2014). Awards of call options on stocks can also induce managers to understate earnings. The stock options are typically granted with a fixed exercise price equal to the stock price on the grant date. The value of call options on stocks is inversely related to the exercise price of the stock options. Therefore, managers have an incentive to deflate earnings prior to option grants to lower the exercise prices of the stock options (Baker et al. 2003; Bartov and Mohanram 2004). In line with this notion, Coles et al. (2006) find evidence of downwards earnings management in the periods following announcements of cancellations of executive stock options up to the time the options are reissued.

In essence, to the extent that investors rely on reported earnings in forming beliefs about future earnings and to trade stocks (Stein 1989), stock prices are likely to be distorted by earnings management, thereby fulfilling various foregoing opportunistic incentives of managers.

88.2 Various Tactics of Earnings Management via Accruals Manipulation

To make earnings well capture a firm's real performance, accounting standards allow firm management to have discretion in making accounting policies and estimates on various noncash business transactions/items, such as allowance for doubtful receivables, impairment losses on inventories, depreciation of tangible assets, which are often uncertain in terms of their actual amount accruing.¹ Such choices of accounting policies and estimates by managers result in accruals, which are part of earnings. Accruals are largely subject to managers' estimation of the nature and amount of a firm's noncash business transactions/events, and therefore can be manipulated by managers for the purpose of distorting reported earnings.

Below we set forth the methods, which we believe are common in practice, of manipulating accruals to inflate earnings. First is the understatement of depreciation (amortization) expense for long-term (in)angible assets. Accounting rules require that firms recognize depreciation or amortization expense on an annual basis for the reduction in the book value of long-term assets. Upon acquiring a long-term asset, managers may overestimate its economic useful life or overestimate its salvage value, such that depreciation or amortization expense for the asset will be understated. Managers may also change the depreciation method from the straight-line (or accelerated) method to the deceleration (or straight-line) method, such that the depreciation or amortization expense at the early stage of the asset's economic useful life will be understated.

The second common tactic is the capitalization of expenditures as assets recognized in the balance sheet, which should be expensed according to the related accounting standards; this would lead to inflations of both reported earnings and assets. For instance, when acquiring a long-term asset, managers capitalize costs (e.g., fine for wrong parking of trucks that transport the asset) not necessarily incurred to bring the asset to its intended use. After acquisition of a long-term asset, managers capitalize the expenditures (e.g., those incurred for repairs, maintenance, additions, or litigation costs) even though these outlays are expected to benefit the firm only for the current year. Managers may even capitalize expenditures on research and development, software development layouts, and brands and membership bases that are created through advertising and promotions; the benefits

¹Throughout this chapter, all the accounting standards and policies mentioned are based on the general accepted accounting principle (GAAP).

associated with these expenditures are highly uncertain, hence the expenditures should be expensed rather than capitalized.²

The third common strategy is to delay the write-down of current or long-term assets. Once the value of the assets (e.g., intangible assets, inventories, property, plant, and equipment (PPE)) is impaired, firms are required to recognize impairment losses. Managers might not do so even when the fair value of an asset falls substantively below its book value. When managers over-buy or over-produce products in excess of market demand for the products, the value of inventories is likely to decrease; when PPE fail to sustain the production of products that meet the changing needs, preferences, and tastes of customers, the value of the PPE will decrease for the firm; when a firm is discovered by the public to have been involved in environmental, social, and governance scandals, the firm will suffer from reputational losses, and its goodwill will be impaired as a result. In all these cases, managers should write-down the book values of the assets to their fair values on a timely basis. Delayed write-down of the assets will lead to the overstatement of reported profits and total assets.³

The fourth common maneuver is to accelerate the recognition of revenues. Managers should recognize revenues only when products (services) have been delivered (provided) to customers and when cash collection is reasonably likely. Nonetheless, managers may recognize revenues when they receive cash but have not delivered products, or render services, to customers. Firms that bundle product sales with service contracts are particularly prone to, either intentionally or unintentionally, advance recognition of revenues for the services that have not yet been provided, especially for long-term service contracts. Managers may also make the advance recognition of revenues in cases when invoicing for the product delivery or service rendering has not yet occurred.

The fifth ordinary approach to boost earnings via accruals is the understatement of allowance for doubtful receivables or loan losses. Firms are required to estimate their expected customer defaults on trade receivables or loans. Managers might underestimate the amount of likely default payment by customers. As a consequence, allowance for doubtful accounts (loan losses), which are contra-accounts to receivables (loans) in

²It is worth noting, though, that unlike GAAP, the International Financial Reporting Standards (IFRS) No. IAS 38 stipulate that research and development expenditures that meet certain specified criteria can be capitalized as an intangible asset. Further details can be found via the link: <https://www.ifrs.org/issued-standards/list-of-standards/ias-38-intangible-assets/#about>

³Pursuant to IFRS No. IAS 2, inventories are measured at the lower of costs and net realizable value, the latter of which equals the estimated selling price minus the estimated costs of completion and the estimated costs to sell. Details can be found via the link: <https://www.ifrs.org/issued-standards/list-of-standards/ias-2-inventories/>. According to IFRS No. IAS 36, if the carrying value of noncurrent assets (e.g., goodwill, PPE) exceeds their recoverable value, the assets should be impaired to their recoverable value, which is the lower amount of (i) fair value less costs to sell and (ii) value in use. The impairment losses can be reversed for the assets, except for goodwill, under the guidance of IFRS, whereas the reversion is not allowed if a firm follows GAAP. Further details can be found via the link: <https://www.ifrs.org/issued-standards/list-of-standards/ias-36-impairment-of-assets/>

the balance sheet, and the associated bad-debt-provision expense will be understated, leading to inflation of both reported earnings and total assets.

The sixth common device to inflate accruals and earnings is to underestimate various provisions (e.g., warranty coverage for products sold, provision for legal fees, provision for environmental clean-up fees) or put the provisions off balance sheet. To illustrate, the degree of future expenditures on warranty coverage for products (services) sold (rendered) is often uncertain. Managers may underestimate provisions on warranty coverage and thereby inflate earnings. Accounting standards prescribe that a firm recognizes a provision when the obligation is unavoidable, is likely to lead to a future outflow of cash, and can be reliably estimated. Managers may understate, or even not recognize, the provision, even when a liability meets all the requirements for recognition. In consequence, the associated provision expense will be understated, and earnings will be overstated.

The last common means to prop up earnings via accruals is the understatement of pension and postretirement benefit obligations. Managers are required to estimate and report the present value of the pension benefits and other postretirement benefits that will have been earned by employees over their years of working and services for their firm. These benefits pertain to the firm's liability to its employees, which is offset by any assets (comprised mainly of cash) that the firm set aside in the postretirement plan. The pension and postretirement benefit obligations are estimated based on forecasts of future wages, employee attrition rates, and the expected lives of retirees. Managers can make the forecasts to be pessimistic, such that the firm's pension obligation and associated pension expenses are understated, thereby achieving the end of inflating reported earnings.

Managers normally have the best information on the uncertainties surrounding accruals recognition. The estimation and reporting of accruals involving the foregoing seven common tactics of accruals manipulation are hard to be challenged by outsiders who do not have access to private corporate information and are generally constrained in their abilities to process value-relevant information. This strengthens managerial incentives to make use of the accruals-manipulation tactics to boost reported earnings. By the same token, managers can manage earnings downwards by (i) overstating depreciation for noncurrent assets, (ii) delaying recognition of revenues, (iii) accelerating write-down of noncurrent or current assets, (iv) overestimating pension and postretirement benefit obligation, (v) overestimating allowance for doubtful receivables, loan losses, or inventory obsolescence, and (vi) overestimating various provisions (e.g., warranty coverage for products sold, provision for legal fees, provision for environmental clean-up fees).

88.3 Various Tactics of Earnings Management via Real Activities Manipulation

Real earnings management is the purposeful action of managers to manipulate reported earnings by altering real business activities strategically. Unlike accruals-based earnings management, which changes the accounting methods and/or

estimates, real activities manipulation is achieved by way of manipulating the timing and structuring of real business transactions.

There are three common ways of manipulating recurring business activities to inflate earnings (Roychowdhury 2006). First is to accelerate the timing of sales through increased price discounts or more lenient credit terms to increase sale revenues for the current fiscal year. Second is to report lower costs of goods sold through overproduction during the current fiscal year. Per unit fixed overhead cost would be reduced as a result of overproduction, resulting in a reduction of costs of goods sold, and thus an increase in net income, for the current period. But the costs of unsold goods would increase as a result and be eventually realized into the recognition of higher costs of goods sold in future years. Third is to reduce the current advertising, research and development, maintenance and repair, and other selling, general, and administrative expenditures that would benefit a firm in the long run.

There are other ways of real earnings management, which are through nonrecurring business activities. First, the number of shares outstanding would decrease as a result of stock repurchase, resulting in an increase in earnings per share (EPS). Thus, managers can launch stock repurchases for the purpose of increasing EPS (Hribar et al. 2006; Burnett et al. 2012). Second, managers make advance sales of an asset, which can still create value for a firm, to increase reported profits for the current fiscal year (Bartov 1993; Herrmann et al. 2003). Third, a firm would not have to pay interest expense any longer after the debt is swapped by equity by the firm with its creditors. Thus, managers may advance debt-equity swaps to inflate reported profits, though their firm is not in financial constraints or distress and thus does not have to make the debt-equity swaps for the purpose of relieving the constraints/distress (Hand 1989). Fourth, securitized accounts receivables should be sold at a point when the expected likelihood of default payments by the customers is the lowest, such that the associated securitization expense will be lowest. Managers may delay the sales of securitized accounts receivables to boost reported profits (Dechow and Shakespear 2009). Fifth, managers advance sales of securities that are aimed for long-term investments and for taking control of related business parties.

Last but not least, real earnings management could be done via profit-shifting among related parties inside a business group. For example, real profits can be transferred between a parent company and its subsidiaries, among subsidiaries, or between a firm and its main, non-subsidiary suppliers or customers by means of getting products transacted at a price different from the market price. Or rather, a firm may manage earnings upwards by buying (selling) products from (to) its related business parties at a price lower (higher) than the market price, and may manage earnings downwards by making the purchases (sales) at a price higher (lower) than the market price. Even without transacting products at a price different from the market price, managers may shift income from a parent company to its not-wholly-owned subsidiaries to lower the net income of the consolidated income statement of the parent company, as earnings shared by minority investors of its subsidiaries would not be recognized in the consolidated income statement. Likewise, managers

may shift income from the not-wholly-owned subsidiaries to their parent company to make its consolidated net income higher.

88.4 Consequences and Determinants of Earnings Management

Were managers to inflate (deflate) earnings via accruals in the current year, the inflated (deflated) accruals would be reversed downwards (upwards) in subsequent years, leading to a fall (rise) in earnings and thereby plausible suspicion by outsiders for accruals manipulation (He et al. 2020b). There is no cost to managers for accruals manipulation if it goes undetected by outsiders. Nevertheless, if accruals manipulation is uncovered by the public, the firm will lose reputational capital, defined as the present value of improvements in net cash flows that arise when a firm's stakeholders trust that the firm will uphold and abide by its explicit and implicit contracts and will not act opportunistically to the detriment of stakeholders (Amiram et al. 2018). The misconduct firm would be subject to higher costs of capital, lower sales, or higher operating costs, as the revelation of accruals manipulation changes the terms of contracts or tacit agreements by which varied stakeholders are willing to do business with the firm. The misconduct firm might also be subject to regulatory fines, class-action settlements, and legal expenses. The legal costs rest on the extent of accruals manipulation and the degree of losses it causes to investors and other financial statement users. As a result of the reputational losses and threat of litigation, it will be difficult or costly for the firm to finance and/or contract for investments and operations in the future, thus reducing firm value and increasing corporate default risk (Ball and Shivakumar 2008; He 2015, 2016). There is an array of evidence (e.g., Dechow et al. 1996; Beneish 1999; Burns and Kedia 2006; Karpoff et al. 2008; Gande and Lewis 2009) to suggest that a drop in firm value as a consequence of the discovery of accruals inflation is substantively larger than the original stock price inflation.

The manipulation of accruals might also have adverse impact on the firm's business activities. For instance, companies often make investment decisions in the current period based on their past performance (He et al. 2020a). The inconsistency between the current period's investments and the previous periods' manipulated earnings might induce investors' suspicion of accruals misreporting. To lower the risk of detection, managers may adjust their firm's investments to be in line with the managed earnings. As such, accruals inflation might lead managers to overinvest in risky projects that increase corporate default risk (e.g., McNichols and Stubben 2008; Kedia and Philippon 2009; Kumar and Langberg 2009); in a similar vein, accruals deflation might cause managers to underinvest in positive-net-present-value projects that could guard against financial distress for the firm in the long run (e.g., McNichols and Stubben 2008; Kedia and Philippon 2009; Kumar and Langberg 2009; He 2015). Put differently, managers who *unintentionally* inflate (deflated) reported earnings believe the misreported growth trend and would overinvest (underinvest) accordingly in the financial markets and product markets; those who *intentionally* boost (lower) reported earnings would seek to overinvest (underinvest)

as well to maintain consistency between misreported earnings and investments so as to avoid outsiders' suspicions of earnings management (Sadka 2006; McNichols and Stubben 2008; Kedia and Philippon 2009). In addition, managers might lose their jobs and compensations once accruals manipulation is discovered by the public.

Unlike accruals-based earnings management, real earnings management is based on the manipulation of real activities in ways that deviate from optimal business strategies and normal business practices (Roychowdhury 2006; Zang 2012). Therefore, real activities manipulation, even if not detected by outsiders, will have a profound, negative influence on the firm's future performance. The larger the deviation of manipulated real activities from normal business practices, the more adverse the influence for the firm. If the manipulation is detected by outsiders, the adverse influence will be even stronger, as the detection will bring about reputational costs to the firm just as the discovery of accruals-based earnings management does.

Whether managers will manage earnings, either via real activities manipulation or accruals manipulation, depends fundamentally on whether the benefits from the earnings management are perceived by managers to be higher than the costs. Accordingly, the determinants of earnings management are related to the perceived benefits versus costs of managing earnings. Taking the benefits as given in the foregoing settings in which managers have incentives to manage earnings, we may take the costs as varying on different firm characteristics. Larger firms, higher-growth firms, financially healthier firms, lower-risk firms, better-performing firms, and firms that face fierce industrial product market competition are less likely to manage earnings, as the reputational costs of doing so are higher for these firms. Firms with stronger internal governance, firms with more intense external monitoring, and firms whose financial statements are subject to high-quality audits are less prone to manage earnings, as the risk of detection of earnings management tends to be higher for these firms.

Below we set out a review of literature on the determinants of earnings management.

Firm size: A great deal of research provides evidence to suggest that firm size is negatively associated with the magnitude of earnings management. Large firms are more likely to have good internal control, since they normally have more assets and business transactions that need be monitored well, and have more resources to invest in internal control systems. In contrast, small firms tend to have weaker internal controls and are more likely to manipulate earnings (e.g., Kinney and McDaniel 1989; Ge and McVay 2005; Doyle et al. 2007; Ashbaugh-Skaife et al. 2007; Haw et al. 2004).

Growth prospect: Some prior research finds that high-growth firms have lower earnings persistence (Nissim and Penman 2001; Penman and Zhang 2002), which makes investors vigilant of potential earnings management, and have fewer earnings management opportunities (Richardson et al. 2005), and that high-growth firms have strong internal control in place to curb potential financial misconduct that impedes growth (Doyle et al. 2007; Ashbaugh-Skaife et al. 2007). In low-growth firms which often have surplus free cash flow and associated agency problems, managers tend to

purposefully manage earnings to avoid reporting any loss that results from rent extraction (Chung et al. 2005).⁴

Financial constraints: Financially constrained firms are those that do not have sufficient internal funds for investments and are subject to high costs for accessing capital markets (Korajczyk and Levy 2003). Financially constrained firms have incentives to prop-up reported earnings so as to raise external capital at a lower cost. Consistent with this notion, Campello et al. (2010) find that, when facing tight economic conditions, financially constrained firms tend to boost earnings through real activities such as a reduction in expenditures on technology and marketing. Linck et al. (2013) show that, given valuable investment opportunities, firms that are in financial constraints tend to use accruals to highlight good news.

Financial distress: Financially distressed firms are referred to as those that are unable to fulfil debt obligations in the foreseeable future. If financial distress of the firms is prolonged, they would be likely to go bankrupt. Prior studies find that financially distressed firms are likely to manage earnings upwards to conceal deteriorating performance and to avoid reported losses (Graham et al. 2005; Jaggi and Lee 2002; Saleh and Ahmed 2005; Rosner 2003).

Risk: One motivation for earnings management is to alleviate the market's perception about firm risk. An increase in firm risk could be associated with a higher likelihood and extent of earnings management. We classify firm risk into business risk and information risk, the former of which includes operational risk and financial risk.

Operational risk: Previous research (Rountree et al. 2008; Hunt et al. 1997) shows that operational risk, proxied by earnings volatility or cash flows volatility, is valued negatively by investors. They would charge a higher premium for investing in a firm that has high operational risk. Therefore, managers are likely to manipulate accruals to smooth earnings so as to mitigate any adverse impact of volatile operating performance of their firm and thereby lower its cost of capital (Francis et al. 2004).

Financial risk: Financial risk refers to the risk associated with external financing, and hinges largely on the level of capital structure. Theory (e.g., Modigliani and Miller 1958, 1963; Mayers and Majluf 1984; Desai et al. 2007) suggests that a good capital structure requires a good trade-off between debt and equity in the way a firm is financed. This theory has been examined in the firm-level, cross-sectional empirical analyses, with results revealing that smaller, riskier firms and those with more intangible assets, higher profitability, and stronger growth prospect are more suited for using equity rather than debt to finance their investments (e.g., Hovakimian et al. 2001; Frank and Goyal 2009; Kayhan and Titman 2007; Adam and Goyal 2008). A firm's optimal capital structure can be determined by estimating the fitted value from a regression of financial leverage on its typical determinants (i.e., firm size, risk, asset tangibility, profitability, and growth). Any deviation from the optimal capital

⁴Surplus free cash flows might induce agency problems in which managers use the cash flows for self-serving purposes.

structure might increase the firm's financial risk (Leland 1998). If financially risky firms wish to obtain new loans, the lenders will likely set tight covenants. To avoid debt-covenant violations, the firms are likely to manipulate earnings upwards. This has been documented extensively by the extant literature (e.g., Sweeney 1994; Dichev and Skinner 2002; Beatty and Weber 2003).

Information risk: Bhattacharya et al. (2003) define information risk as the risks that investors may confront due to the insufficient, imprecise, or even biased information on which to base their investment decisions. A large body of studies show that the cost of capital increases with increased information risk (Bhattacharya et al. 2003; Francis et al. 2005; Bhattacharya et al. 2011; Bharath et al. 2008; Qi et al. 2010). These studies use accruals quality as the proxy for information risk, suggesting that accruals-based earnings management would induce high information risk to a firm.

Performance: Firms that are not performing well have incentives to inflate earnings to conceal the unsatisfactory performance (Petroni 1992; Dechow et al. 1995; DeFond and Park 1997; Balsam et al. 1995; Keating and Zimmerman 1999; Doyle et al. 2007; Kinney and McDaniel 1989). On the other hand, the opportunity costs (i.e., personal reputational losses and compensation losses) for managers to engage in earnings management are likely to be higher for better-performing firms, hence they are less likely to get involved in earnings management.

Competition: Prior studies have investigated the relation between industrial product market competition and earnings management, but document mixed views and evidence on the relation. Some researchers argue that competition might incentivize firms to inflate earnings to attract customers and fund providers (Shleifer 2004; Bagnoli and Watts 2010; Markarian and Santalo' 2014; Karuna et al. 2015). However, some argue that industrial competition has a deterrence effect on managers' misconduct (e.g., Giroud and Mueller 2011; Guadalupe and Pérez-González 2010) and earnings management behavior (Balakrishnan and Cohen 2013), because the reputational costs are higher for firms subject to fierce competition. Some studies also find that competition would impact a firm's trade-off between different tactics of earnings management. For instance, Zang (2012) finds that (non-)market leaders in an industry are more likely to resort to real activities (accruals) to inflate earnings. Shi et al. (2018) find evidence that competition is positively (negatively) associated with accruals-based (real) earnings management.

External monitoring: Extensive evidence suggests that external monitoring from auditors, financial analysts, and institutional investors mitigates earnings management. Auditing is an important external governance mechanism, which provides assurance of earnings quality to external stakeholders by monitoring the financial reporting process and curbing any misreporting therein. The existing literature (e.g., Becker et al. 1998; Francis et al. 1999; Balsam et al. 2003; Krishnan 2003) documents that high-quality auditing reduces accruals-based earnings management. However, real earnings management is less likely to be detected by auditors, thus managers tend to engage in real activities to inflate earnings when the accruals manipulation is constrained by higher-quality audits (Chi et al. 2011).

Financial analysts and institutional investors tend to monitor earnings management because they need to rely on high-quality financial reports for purposes of valuation and/or performance evaluation. Analysts generally have the training and industrial background in accounting and finance, and are thus advantageous in monitoring via collection and analysis of various firm-specific information and in uncovering accounting distortions (e.g., Jensen and Meckling 1976; Healy and Palepu 2001; He et al. 2019a). Miller (2006) and Dyck et al. (2010) offer evidence to suggest that analysts play a critical role in detecting corporate frauds in the financial marketplace. Yu (2008) and Degeorge et al. (2013) show that earnings management is negatively associated with analyst coverage. As to the role institutional investors play in monitoring earnings management, prior research shows that institutional stock ownership is negatively associated with both accruals-based earnings management (Chung et al. 2002; Mitra and Cready 2005) and real earnings management (Bushee 1998; Bange and De Bondt 1998).

Corporate governance: Corporate governance helps reduce the agency problem between shareholders and managers through internal monitoring mechanisms such as board governance, ownership structure, and various governance-related committees (Jensen and Meckling 1976). Prior studies find that firms with higher board independence (e.g., Jensen and Meckling 1976; Dechow and Dichev 2002), dispersed stock ownership structure (e.g., Leuz et al. 2003; Nenova 2003; Dyck and Zingales 2004), and effective internal audit committees (e.g., Carcello et al. 2006; Bedard and Johnstone 2004) have less opportunistic earnings management.

Executive compensation: As noted previously in Sect. 88.1, one incentive for earnings management arises from stock and stock option holdings by executives. A vast literature provides evidence that executive equity compensation is positively associated with earnings management (e.g., Larcker et al. 2007; Bergstresser and Philippon 2006). He (2015) finds that CEOs holding large inside debt are less likely to manipulate accruals. Kohlbeck and Luo (2019) extend He (2015) by showing that CFO inside debt is negatively associated with accruals-based earnings management.

There are various empirical measures of the determinants of earnings management that are documented in prior research. We summarize these measures in Table 1.

88.5 Trade-Off Between Accruals Manipulation and Real Activities Manipulation to Manage Earnings

As noted earlier, whether managers are inclined to manage earnings hinges crucially on whether the expected benefits of doing so outweigh the expected costs. Since the costs of managing accruals differ from those of managing real business activities, managers will have to trade-off between accruals manipulation and real activities manipulation to manage earnings. Manipulating earnings via accruals does not affect the underlying business activities and cash flows of a firm, whereas managing earnings via real business activities does, and thereof will impose negative economic

Table 1 Measurements of the determinants of earnings management

Determinants	Measurements
Firm size	The natural logarithm of the market or book value of total assets or equity
Growth prospect	The book-to-market ratio; sales growth; assets growth; the ratio of capital expenditures to sales; the price-to-book (P/B) ratio
Financial constraints	SA index (Hadlock and Pierce 2010); WW index (Whited and Wu 2006); KZ index (Kaplan and Zingales 1997); dividend payout ratio (Denis and Sibikov 2010; He et al. 2021); retained earnings scaled by total equity; cash and cash equivalents scaled by total assets; cash flows from operating activities scaled by total assets
Financial distress	Distance-to-default and expected default frequency (Campbell et al. 2008); Altman Z-score (Altman 1968); liquidity ratios (e.g., current ratio, quick ratio, cash ratio); solvency ratios (e.g., financial leverage, interest coverage ratio; cash coverage ratio)
Business risk	The standard deviation of earnings, sales, stock returns, or cash flows from operating and investing activities; the standard deviation of return on assets (ROA), return on equity (ROE), or return on sales (ROS)
Performance	Return on assets (ROA); return on equity (ROE); return on sales (ROS); gross profit margin ratio; stock returns; the natural logarithm of sales; Tobin's Q (Kaldor 1966); total factor productivity (Sickles and Zelenyuk 2019)
Competition	Herfindahl-Hirschman index on sales of varied firms in an industry (Harris 1998); product substitutability, market size of competing products, entry costs (Karuna 2007; He 2018; He et al. 2019b)
External monitor	Analyst coverage; institutional stock ownership; audit quality (the presence of big-4 audits, the presence of internal audit committee, the ratio of audit fees to sales revenues of a firm, auditor tenure, auditor switches; auditor industrial specialization; total audit revenues of an audit firm)
Corporate governance	Board independence (e.g., the portion of independent outside directors on the board); CEO power (e.g., whether CEO is also the chairman of the board); ownership structure (e.g., Herfindahl-Hirschman index on shares of different shareholders in a company)
Executive compensation	Executive equity compensation (stocks and stock options of executives scaled by total shares outstanding; equity delta; equity vega); executive debt-like compensation (pension and deferred compensation of executives scaled by total assets)

consequences on a firm. Nonetheless, accruals manipulation is subject to scrutiny from regulators and auditors, and entails a risk of detection and litigation by financial statement users. In contrast, real earnings management, which departs a firm from normal business practices without altering the accounting methods and/or estimates, entails relatively lower detection risk and is not subject to any legal discipline or legal cost. Therefore, given a desired level of earnings, managers tend to use accruals-based earnings management and real earnings management as substitutive instruments to manipulate earnings (e.g., Cohen et al. 2008; Bhojraj et al. 2009; Zang 2012; Cohen and Zarowin 2010; Badertscher 2011; Gunny 2010; Burnett et al. 2012; Chan et al. 2014; Kothari et al. 2016).

Accounting standards need to trade-off between granting financial-reporting discretion, which leaves room for accruals-based earnings management, and imposing mechanical, rigid reporting criteria, which creates opportunities and incentives for real earnings management. Such a trade-off seeking a good balance between the discretion *vis-à-vis* rigidity should lead to a minimization of latitude for overall earnings management, which is a challenging yet pressing task for accounting-standard setters nowadays.

88.6 How to Discern and Measure Earnings Management?

Earnings consist basically of cash and accruals. The reported amount of cash cannot be altered without changing real business activities, but the reported numbers of accruals can, simply by changing the associated estimations such as allowance for trades receivables, impairment losses on assets, pension obligations. Therefore, one way of assessing earnings quality is to estimate abnormal levels of accruals. Before we introduce the approaches of estimating abnormal accruals, we first expatiate on the definitions and measurements of varied accruals.

Accruals are formally defined as involving changes in noncash assets or noncash liabilities of a firm, which denote its future cash inflows or outflows. Positive (negative) accruals reflect future benefits (obligations), or future cash inflows (outflows), for a firm. Aggregate change in noncash assets net of aggregate change in noncash liabilities represents the net change in expected future benefits. We thus refer to this amount as comprehensive accruals, which we calculate as change in common stockholders' equity less change in cash and cash equivalents. Comprehensive accruals are composed of operating accruals (i.e., accruals relating to operational and investing activities) and financial accruals (i.e., accruals relating to financing activities such as investments in marketable securities, interest-bearing lending or borrowing). Specifically, financial accruals are defined as change in financial assets (e.g., cash, security investments, lending) net of change in financial liabilities (e.g., debt) and of change in cash. All assets other than financial assets, and all liabilities other than financial liabilities, relate to operations of a firm (which are, in a broad sense, inclusive of investment activities).⁵ Accordingly, operating accruals are computed as change in net operating assets exclusive of cash and cash equivalents, or as net operating profits after taxes and before interests less free cash flow. Operating accruals can be further categorized into working capital accruals (e.g., inventories, accounts receivables, accounts payables, which are short-term in nature and expected to reverse, and result in cash inflows or outflows, within a year) and long-term operating accruals (e.g., property, plant, and equipment (PPE), pension obligations, which reflect long-term future benefits or obligations). Working capital accruals are computed as: (change in current assets – change in cash and short-term

⁵Throughout this chapter, we take operating activities as, in broad sense, including investing activities. As such, operating accruals are in themselves inclusive of investing accruals.

investments) – (change in current liabilities – change in debt included in current liabilities); long-term operating accruals are then calculated as operating accruals minus the working capital accruals.

Apart from categorizing accruals into working capital accruals, long-term operating accruals, and financial accruals (Richardson et al. 2005), we may also decompose accruals based on whether they articulate across both balance sheet and statement of cash flows. All the changes in assets or liabilities in balance sheet, which are caused by regular operating transactions, should eventually go through statement of cash flows either in the current year or in future years. However, some “non-articulating” events, such as noncash acquisitions and divestitures, foreign currency translation adjustments, lead to changes in the balance sheet but do not cause any change in the statement of cash flows in the current or future years. As such, the non-articulating events will bring about a difference between accruals measured using the balance sheet approach (i.e., taking changes in noncash assets or liabilities in the balance sheet) and accruals measured using the cash flow approach (i.e., subtracting operating and investing cash flows and accrued interests from net operating income before interests and after taxes). Such a difference in accruals arises from the “non-articulating” events and is thus termed non-articulating accruals. The rest of accruals are articulating accruals, which are calculated, under the cash flow approach, as net operating income before interests and after taxes less operating and investing cash flows and accrued interests.

To estimate the abnormal level of operating accruals, we run the following modified Jones regression model by industries and years, with at least 20 firms required for each year and industry (Dechow et al. 1995):

$$\frac{OPACC_{i,t}}{Ass_{i,t-1}} = \alpha_0 + \frac{\alpha_1(\Delta REV_{i,t} - \Delta REC_{i,t})}{Ass_{i,t-1}} + \alpha_2 \left(\frac{PPE_{i,t}}{Ass_{i,t-1}} \right) + \varepsilon_{i,t} \quad (1)$$

where $OPACC_{i,t}$ = operating accruals of firm i at year t ; $\Delta REV_{i,t}$ = change in revenues of firm i for year t ; $\Delta REC_{i,t}$ = change in accounts receivable of firm i at year t ; $PPE_{i,t}$ = the net value of property, plant, and equipment of firm i at year t ; and $Ass_{i,t-1}$ = total assets of firm i at year $t-1$. Under normal operating circumstances, higher growth in cash revenues and less PPE (denoting lower depreciation expense) are associated with more operating accruals. Therefore, it is expected that α_1 (α_2) is positive (negative) and statistically significant at a conventional level.⁶ Abnormal accruals of firm i at year t would be the residual value (i.e., the actual value of $\frac{OPACC_{i,t}}{Ass_{i,t-1}}$ less the predicted value of $\frac{OPACC_{i,t}}{Ass_{i,t-1}}$) estimated from the regression model. Significantly positive (negative) abnormal accruals denote upwards (downwards) earnings management by a firm. The assumption underlying the model used to measure the

⁶The parameter estimation for Model (1) (as well as the follow-up Models (2, 3, 4, 5, 6, and 7) used to estimate various types of earnings management) may incorporate a scaled intercept to mitigate potential heteroskedasticity attributed to the variation in the scaling variable (i.e., total assets) (e.g., Kothari et al. 2005; He 2016).

degree of accruals manipulation is that accruals are innately determined by operational fundamentals regarding cash-revenue growth and fixed assets, and that any deviation from such fundamentals is attributed to managerial accruals manipulation. In the presence of non-articulating events (typically, mergers and acquisitions) for a firm, operating accruals (*OPACC*) should be replaced by articulating operating accruals (*A_ACC*) in Model (1) to estimate the abnormal articulating operating accruals.

Working capital accruals can be manipulated by managers at any year they wish, but the manipulated accruals will be likely reversed in the next year, exposing a firm to a relatively higher risk of detection by outside investors. In contrast, long-term operating accruals cannot be manipulated at any year in which managers wish for a higher or lower level of earnings. For instance, at year *t*, managers wish for inflated earnings. But the market values of operating fixed assets are all higher than the book values of the assets, hence there is no way for managers to delay or underestimate the asset write-off in order to inflate earnings. Alternatively, managers can time their opportunistic incentives for inflating earnings to be at the point when they can manipulate the long-term operating accruals upwards. Therefore, compared with working capital accruals, long-term operating accruals are less flexible for managers to manipulate in order to achieve opportunistic objectives. That said, any manipulated long-term operating accruals tend to reverse in the distant future, and thus are less likely to be detected by outsiders. If we seek to assess whether and to what extent managers use working capital accruals (*WCACC*) to manipulate earnings, then Model (1) should be replaced with the following model to estimate abnormal working capital accruals:

$$\frac{WCACC_{i,t}}{Ass_{i,t-1}} = \alpha_0 + \frac{\alpha_1(\Delta REV_{i,t} - \Delta REC_{i,t})}{Ass_{i,t-1}} + \varepsilon_{i,t} \quad (2)$$

If we wish to assess whether and to what degree managers use long-term operating accruals (*LTACC*) to manage earnings, then Model (1) should be replaced with the following model to estimate abnormal long-term operating accruals:

$$\frac{LTACC_{i,t}}{Ass_{i,t-1}} = \alpha_0 + \alpha_1 \left(\frac{PPE_{i,t}}{Ass_{i,t-1}} \right) + \varepsilon_{i,t} \quad (3)$$

It is worth noting (i) that the abnormal accruals estimated from Model (1) might be attributed to the differences in accounting methods adopted by firms in or between different industries, rather than to managerial accruals manipulation and (ii) that cash flows from operations (*CFO*) in the past, current, and future years are also potential innate determinants of operating accruals for the current year. To account for these concerns in the assessment of accruals manipulation by firms, Francis et al. (2005) develop the following modified Dechow and Dichev model, which is estimated by industries and years and with at least 20 firms required of the sample for each year and industry:

$$\begin{aligned} \frac{OPACC_{i,t}}{Ass_{i,t-1}} = & \alpha_0 + \alpha_1 \left(\frac{CFO_{i,t+1}}{Ass_{i,t-1}} \right) + \alpha_2 \left(\frac{CFO_{i,t}}{Ass_{i,t-1}} \right) + \alpha_3 \left(\frac{CFO_{i,t-1}}{Ass_{i,t-1}} \right) \\ & + \frac{\alpha_4(\Delta REV_{i,t} - \Delta REC_{i,t})}{Ass_{i,t-1}} + \alpha_5 \left(\frac{PPE_{i,t}}{Ass_{i,t-1}} \right) + \varepsilon_{i,t} \end{aligned} \quad (4)$$

The standard deviation of the residual values estimated from Model (4) for years t-4 to t is used as the refined measure of a firm's accruals manipulation. The premises underlying such a measure are that (i) any measurement error in determining accruals is attributable to either noise or management errors and (ii) that noise is time-invariant from year t-4 to year t. Operating accruals and operating cash flows tend to be positively correlated across different years in a time-series but negatively correlated for a given year in the cross-section. Therefore, α_1 and α_3 should be positive and statistically significant at conventional levels, while α_2 should be significantly negative. High-quality accruals should well reflect future, current, and past operating cash flows, after controlling for PPE and changes in cash revenues. As with Model (1), α_4 (α_5) should be significantly positive (negative).

As mentioned earlier, three common ways of manipulating recurring business activities to inflate earnings are (i) accelerating the timing of sales through increased price discounts or more lenient credit terms, (ii) lowering costs of goods sold through overproduction, and (iii) reducing advertising, research and development, maintenance and repair, and other SG&A expenditures. Price discounts and grants of lenient credit terms would result in abnormally low cash flows from operations (CFO). To estimate the resulting abnormal CFO, we run the following regression model by industries and years, with at least 20 firms required of the sample for each year and industry:

$$\begin{aligned} \frac{CFO_{i,t}}{Ass_{i,t-1}} = & \alpha_0 + \alpha_1 \left(\frac{Sales_{i,t}}{Ass_{i,t-1}} \right) + \alpha_2 \left(\frac{\Delta Sales_{i,t}}{Ass_{i,t-1}} \right) + \alpha_3 \left(\frac{Prod_{i,t}}{Ass_{i,t-1}} \right) \\ & + \alpha_4 \left(\frac{Disexp_{i,t}}{Ass_{i,t-1}} \right) + \varepsilon_{i,t} \end{aligned} \quad (5)$$

where $CFO_{i,t}$ = cash flows from operations of firm i in year t; $Sales_{i,t}$ = sales revenues of firm i in year t; $Ass_{i,t-1}$ = total assets of firm i in year t-1; $Prod_{i,t}$ = costs of goods sold + change in inventories; and $Disexp_{i,t-1}$ = advertising expenses + research and development expenses + selling, general, and administrative expenses of firm i in year t. As with Roychowdhury (2006) and Cohen and Zarowin (2010), we control for sales and change in sales for the current period, as they are the innate determinants of the current period's CFO, with larger sales and higher sales growth associated contemporaneously with more CFO. Any deviation from the two determinants is attributed to real activities manipulation by managers. Because real earnings management in terms of overproduction (reduction of discretionary expenses) has a negative (positive) impact on contemporaneous abnormal CFO too, we also include production costs and discretionary expense in Model (5). It is

expected that α_1 and α_2 are positive and statistically significant at conventional levels, and that α_3 and α_4 are significantly negative. The abnormal CFO attributed to price discounts and/or grants of lenient credit terms is measured by the residual value (i.e., the actual value of $\frac{CFO_{i,t}}{Ass_{i,t-1}}$ minus the predicted value of $\frac{CFO_{i,t}}{Ass_{i,t-1}}$) estimated from Model (5). The abnormal CFO being significantly negative denotes upwards real earnings management in respect of price discounts and/or grants of lenient credit terms.

Real earnings management in terms of overproduction would lead to abnormally high production costs. To estimate the abnormal production costs, we run the following regression model by industries and years, with at least 20 firms required of the sample for each year and industry (Dechow et al. 1998; Roychowdhury 2006):

$$\frac{Prod_{i,t}}{Ass_{i,t-1}} = \alpha_0 + \alpha_1 \left(\frac{Sales_{i,t}}{Ass_{i,t-1}} \right) + \alpha_2 \left(\frac{\Delta Sales_{i,t}}{Ass_{i,t-1}} \right) + \alpha_3 \left(\frac{\Delta Sales_{i,t-1}}{Ass_{i,t-1}} \right) + \varepsilon_{i,t} \quad (6)$$

Abnormal production costs of firm i at year t are equal to the residual value (i.e., the actual value of $\frac{Prod_{i,t}}{Ass_{i,t-1}}$ minus the predicted value of $\frac{Prod_{i,t}}{Ass_{i,t-1}}$ estimated from the regression model). Production costs comprise costs of goods sold and change in inventories, with the former (latter) positively associated with the current year's sales (with the current and previous years' sales growth) (Dechow et al. 1998; Roychowdhury 2006). Accordingly, we expect α_1 , α_2 , and α_3 to be positive and statistically significant at conventional levels. The rationale underlying this modeling is that production costs are innately determined by the sales and sales growth, and that any deviation from the innate determinants is attributed to real earnings management as to overproduction. Significantly positive abnormal production costs denote upwards real earnings management via overproduction by a firm.

Real earnings management via strategic reduction in discretionary expenditures would result in abnormally low discretionary expenses. To estimate the resulting abnormal expenses, we run the following regression model by industries and years, with at least 20 firms required of the sample for each year and industry (Dechow et al. 1998; Roychowdhury 2006):

$$\frac{Disexp_{i,t}}{Ass_{i,t-1}} = \alpha_0 + \alpha_1 \left(\frac{Sales_{i,t-1}}{Ass_{i,t-1}} \right) + \varepsilon_{i,t} \quad (7)$$

Abnormal discretionary expenses of firm i at year t are equal to the residual value (i.e., the actual value of $\frac{Disexp_{i,t}}{Ass_{i,t-1}}$ minus the predicted value of $\frac{Disexp_{i,t}}{Ass_{i,t-1}}$) estimated from Model (7). The current period's advertising expenses, research and development expenses, selling, general, and administrative expenses are incurred largely based on the previous period's sales revenues. The higher the sales for the previous year, the high discretionary expenses incurred for the current year. Therefore, we expect α_1 to be positive and statistically significant at a conventional level. A significantly downward deviation from the innate determinant of discretionary expenses (i.e., sales lagged for 1 year) is attributed to managers strategically cutting down on discretionary expenditures for the purpose of propping up reported earnings.

Lastly, we may aggregate the three real earnings management measures into one proxy either by subtracting the abnormal discretionary expenses and the abnormal CFO from the abnormal production costs, or by using common factor analysis to construct a composite measure.

In addition to the foregoing regression analysis, we may use various red flags regarding abnormal accounting changes to appraise potential earnings management by a firm. We summarize a host of red flags, which can be commonly used in practice, in Table 2.

It should be noted that the red flags might be merely reflective of a firm's changed business strategies and/or circumstances, rather than attributed to earnings management motives. For example, a firm which has abnormal increases in its inventories relative to sales might have been introducing a new, better product or product line, rather than overproducing the existing type of products. Abnormal decreases in the allowance for uncollectible receivables relative to sales might be due to a firm changing customer base in ways that cover more financially healthy customers, rather than the firm understating its allowance for uncollectible receivables. Thus, it is important to appraise a firm's potential red flags in the context of its business strategies and/or circumstances. When a firm has high business risk, it will be difficult for us to distinguish whether red flags are attributed to unexpected changes in a firm's business conditions or to questionable accounting numbers. In such a case, we had better use the foregoing regression analysis rather than the red flags to assess the firm's accounting quality.

88.7 Introduction of Various Equity-Valuation Models

The intrinsic value of a firm or stock is determined based on the present value of expected future payoffs from the firm or stock to an investor. The future expected payoffs are discounted at a rate reflecting the time value of money and the uncertainty of these payoffs. The stock price which represents the market expectation of a firm's intrinsic value does not necessarily equal the value of a stock at any point in time. Therefore, analyzing whether and to what degree the stock price deviates from its intrinsic value provides a reliable basis for investors and financial analysts to assess the investment potential of a firm or stock. Investors could gain by buying (selling) the stock if its price is lower (higher) than its intrinsic value. However, unlike the observable price, the value of a firm is invisible and can only be estimated. To this end, we need to first identify what kind of information is relevant to, and has implications for, the present value of the expected future payoffs, and then utilize the value-relevant information to forecast the future payoffs from the firm. Depending on how the payoffs are measured, four primary equity valuation models can be used: (i) dividends valuation model (DVM) which requires forecasts of future dividends (e.g., Gordon 1962), (ii) residual earnings valuation model (or abnormal earnings valuation model (AEVM)) which requires forecasts of future residual earnings (e.g., Edwards and Bell 1961; Peasnell 1982), (iii) free cash flows valuation model (FCFVM) which requires forecasts of future free cash flows (e.g., Damodaran

Table 2 Red flags on different types of earnings management

Types of earnings management	Red flags
Accelerated recognition of revenues	An abnormal increase in credit sales or trades receivable towards the end of the year
Understatements of allowance for doubtful receivables or loan losses	An unusual increase in trades receivables relative to sales increases; significantly growing days' receivable; business downturns for a firm's major customers; significantly growing loan delinquencies
Understatements of pension and postretirement obligations	An abnormally large or increasing pension deficit (i.e., the case in which the funds set aside in the pension benefit plans are significantly less than the pension obligations)
Capitalization of expenditures which should otherwise be expensed	A substantive decline in relevant operating asset turnover ratio
Delayed write-down of current operating assets such as inventories	An unusual increase in inventories relative to sales increases; significantly growing days' inventory; significant write-down of inventories by noncompetitors in the same industry; business downturns for a firm's major clients
Delayed write-down of long-term operating assets	A substantive decline in the long-term operating asset turnover ratio; declines in return on assets (return on equity) to a level lower than the weighted average cost of capital (cost of equity capital); significant write-down by noncompetitors in the same industry which have also suffered from deterioration of long-term operating assets; overpayments for, or unsuccessful, mergers and acquisitions
Understatements of depreciation or amortization on long-term operating assets	An unusual decrease in the overall depreciation (amortization) rate of long-term (in)tangible operating assets; a substantive decline in the long-term operating assets turnover ratio
An inflation of overall operating accruals	Operating accruals of a firm is abnormally higher than the industrial average/median of operating accruals; an abnormal increase in the gap between a firm's reported earnings (operating income) and its cash (cash flows from operations)
Acceleration of the timing of sales in terms of increased price discounts or more lenient credit terms	A significant reduction in the gross profit margin ratio; an abnormal increase in trades receivables relative to sales
Reduction of costs of goods sold through overproduction	Production costs of a firm are abnormally higher than the industrial average/median of production costs; an abnormal increase in production costs
Reduction of discretionary expenses (e.g., advertising, research and development, maintenance and repair, and other SG&A expenditures) that would benefit a firm in the long run	Discretionary expenses of a firm are abnormally lower than the industrial average/median of discretionary expenses; an abnormal decrease in discretionary expenses

2002), and (iv) residual operating income valuation model (ROIVM) (or abnormal operating profits valuation model (AOPVM)) that requires forecasts of future abnormal operating profits. The forecasts involving the application of the four valuation models are based on the value-relevant information, which comes primarily from the audited financial statements.

Dividends are the primitive form of shareholder payoffs, as they represent the income received by shareholders for holding a firm's stocks. Nonetheless, to maintain a clean surplus relationship (i.e., change in book value of equity for shareholders equals earnings after dividends) for accurate valuation of the shareholders' equity, the dividends used in the valuation model encompass all cash flows distributed from the firm to shareholders through not only regular dividend payments but also liquidating dividends, stock repurchases, and the stock issuances in which cash collected from shareholders are considered as negative dividends. In applying the dividend valuation model to value a firm's stocks, we discount all the expected future dividends to their present values over the expected life of the firm in the following way (e.g., Gordon 1962):

$$V_0 = \frac{D_1}{(1 + \rho_E)} + \frac{D_2}{(1 + \rho_E)^2} + \dots = \sum_{t=1}^{\infty} \frac{D_t}{(1 + \rho_E)^t} \quad (8)$$

where V_0 is the intrinsic value of a firm or stock, which refers to the present value of expected future payoffs measured by net dividends; and D_t refers to net dividends, computed as cash dividends minus cash from share issuances plus cash from share repurchases, and representing overall net cash flows to shareholders. D_t and V_0 can be either net dividends per share and equity value per share, respectively, or net dividends and equity value, respectively. ρ_E is the cost of equity, which is computed as: $r_f + (r_m - r_f) * \beta$, where r_f is the risk-free rate, $r_f - r_m$ is the market risk premium, and β is the firm's beta. Before being used as the discount rate for the model, ρ_E needs to be adjusted for the effect of any potential accruals-based earnings management (denoting high information risk), of any significant change in the firm's liquidity and solvency positions, and of any substantial change in operational risks of the firm, in the current period as well as the foreseeable future. This adjustment also applies to the discount rate used for the other three valuation models.

Dividend valuation model measures shareholder payoffs by the cash distribution in the form of dividends paid by the firm to shareholders. Free cash flows (i.e., cash flows that are available for distribution to shareholders and debtholders after capital expenditures on investments and operations) can also be used to measure the expected shareholder payoffs for the valuation of a firm or stock. In such a case, the free cash flow valuation model should be used, with the firm's value estimated as (e.g., Damodaran 2002):

$$V_0 = \frac{FCF_1}{(1 + WACC)} + \frac{FCF_2}{(1 + WACC)^2} + \dots = \sum_{t=1}^{\infty} \frac{FCF_t}{(1 + WACC)^t} \quad (9)$$

where V_0 is the intrinsic value of net operating assets, including equity and debt, of the firm; and FCF is free cash flows, calculated as cash flows from short-term

operations before interests and after taxes (CFO) minus capital investments (CI). Free cash flows can also be calculated as net operating profits after taxes and before interests (NOPAT) minus change in net operating assets (ΔNOA), and in specific as: [net income + (after-tax interest expense – after-tax interest income)] – { $\Delta\text{working capital} + [(\text{expenditures for noncurrent operating assets} - \text{sales of noncurrent operating assets}) - \text{depreciation and amortization expense}]$ }. Since free cash flows are for distribution to both shareholders and debtholders, the discount rate we use to divide the free cash flows is the weighted average of the cost of equity and cost of debt, expressed as follows:

$$WACC = \frac{\rho_E * E + \rho_D * (1 - T_c) * D}{E + D} \quad (10)$$

where ρ_E is the cost of equity; ρ_D is the cost of debt which is calculated as net interest expense before taxes (NIE) divided by net debt (ND) (also named net financial obligation). If net financial assets are larger than net financial obligation, or if interest income is larger than interest expense, NIE/ND of an industrial peer firm that has the same credit rating as the firm being valued could be used as the alternative proxy for the firm's expected cost of debt (ρ_D); T_c is corporate income tax rate; E is the book value of equity; and D is the book value of debt.

Earnings are a summary measure of firm performance and thus form an essential basis for investors to make investment decisions (e.g., Biddle et al. 1995; Francis et al. 2003; Graham et al. 2005). It is thus reasonable for valuers to use future earnings as a proxy for the future payoff that an investor expects to get from the firm. Accordingly, we have the residual earnings valuation model (e.g., Edwards and Bell 1961; Peasnell 1982), in which the value of a firm is estimated based on the future residual earnings as follows:

$$V_0 = B_0 + \frac{AE_1}{(1 + \rho_E)} + \frac{AE_2}{(1 + \rho_E)^2} + \dots = B_0 + \sum_{t=1}^{\infty} \frac{AE_t}{(1 + \rho_E)^t} \quad (11)$$

where ρ_E is the cost of equity; B_0 is the book value of common shareholders' equity at the point of doing the equity valuation; V_0 is the intrinsic value of a firm's equity and represents the present value of expected future shareholder payoffs measured by residual earnings; and AE_t is the residual earnings, also known as residual income or abnormal earnings, at year t, and is computed as: $E_t - \rho_E * B_{t-1}$. An alternative way of writing AE and Model (11) is (e.g., Ohlson 1995; Lee et al. 1999):

$$V_0 = B_0 + \sum_{t=1}^{\infty} \frac{(ROE_t - \rho_E) * B_{t-1}}{(1 + \rho_E)^t} \quad (12)$$

where $ROE_t = \frac{E_t}{B_t}$ is the return on equity, representing the expected return on the book value of equity in year t. Model (12) suggests two key value drivers, which are return on equity (ROE) and growth in equity (as reflected in B_t relative to B_{t-1}); when the

expected return on equity is larger (smaller) than the cost of equity, the firm will experience value-added (–destroying) growth.

The residual earnings valuation model values a firm on the basis of the book value of equity with any excess of required earnings added to the book value, and emphasizes the excess earnings created from the equity. It is worth noting that financial activities do not create value for a firm, if financial assets are traded at a fair price, and if this trading does not pertain to a firm's main business. Therefore, it is important to distinguish the operating activities (inclusive of investing activities) from financing activities for valuation purposes. Given that value is created primarily from operating activities, it is reasonable to measure the expected payoffs from a firm by the residual income from net operating assets, namely, residual operating income. Accordingly, we have the residual operating income valuation model expressed as:

$$V_0 = NOA_0 + \sum_{t=1}^{\infty} \frac{AOP_t}{(1 + WACC)^t} \quad (13)$$

where V_0 is the intrinsic value of net operating assets, including equity and debt, of the firm; NOA_0 is the net operating assets at the point of doing the asset valuation; WACC is the weighted average cost of capital, defined as previously; and AOP is the abnormal operating income, also known as residual operating income. It is defined as: $AOP_t = NOPAT_t - WACC * NOA_{t-1}$, where NOPAT refers to the net operating profits after taxes and before interests, and equals net income plus after-taxes net interest expenses. An alternative way of writing AOP and Model (13) is as follows:

$$V_0 = NOA_0 + \sum_{t=1}^{\infty} \frac{(RNOA_t - WACC) * NOA_{t-1}}{(1 + WACC)^t} \quad (14)$$

where $RNOA_t = \frac{NOPAT_t}{NOA_{t-1}}$, representing the expected return on net operating assets at year t . A firm will add (lose) value in year t in which the expected return on net operating assets is larger (smaller) than the weighted average cost of capital.

The dividends valuation model and residual earnings valuation model value the equity of a firm, while the free cash flows valuation model and residual operating income valuation model value the assets of a firm. So, when using the latter two models, the estimated value of net operating assets needs to be further deducted by the market value of net financial obligation, which approximates its book value due to the fixed interest income/expenses stipulated in the debt contracts, to get the value of the firm's equity.

Equity valuation is done by way of estimating the present value of expected future payoffs from investing in the stocks of a firm. Which of the four valuation models will be the best for use in the valuation depends critically on which form of future payoffs is best-founded yet easiest to forecast. Future dividend payouts depend on future cash adequacies, future investment opportunities, future operational needs, and future debt repayments. Future free cash flows hinge on future cash adequacies,

future investment opportunities, and future operational needs. Forecasts of future residual earnings involve forecasts of future financial obligations such as future interest income or expenses and future financial leverage. Future abnormal operating profits rest exclusively with the future operational activities and performance, from which firm value is created. Therefore, the residual operating income valuation model offers the most reliable and scientific ground on which to estimate the intrinsic value of a firm in a relatively accurate manner, compared with the other three valuation models. Below we set forth more details on the application of the residual operating income valuation model.

The residual operating income valuation model anchors the expected shareholders-and-debtholders payoffs on future residual operating income that is fundamentally determined by three main factors. First is a firm's competitive advantage established via differentiation or cost-leadership strategies and against industrial competitive forces or threats. Second is the industrial growth rate which reflects the growing market demand for a firm's products or services. Third is the rate of return required by capital providers, which will be reduced if the uncertainty of future payoffs in terms of business risk, information risk, or market risk decreases. AOP in year t can be expressed in an alternative way as: $AOP_t = NOPAT_t - WACC * NOA_{t-1} = NOPM_t * sales_t - WACC * \frac{sales_t}{ATO_{t-1}}$, where NOPM is the net operating profit margin ratio, and ATO is the operating asset turnover ratio. As such, the forecast of future AOP can be transformed into a job of the forecasts of future sales, NOPM, ATO, and WACC, which are grounded on analyses of a firm's sales performance, operational efficiency, investment efficiency, and financing, respectively.

To facilitate accurate valuation under the residual operating income valuation model, we typically take three steps to forecast future AOP. First, we choose a finite forecast horizon (usually, 5–10 years) within which the specific trends of future sales, NOPM, and ATO can be reasonably forecasted. Second, we forecast future sales, NOPM, and ATO at the chosen forecast horizon. These forecasts should be made based on (i) the analyses of a firm's business strategies and industrial competitive forces, (ii) time-series and cross-sectional analyses of historical sales, NOPM, and ATO, and (iii) real earnings management, if any. Third, we make some simple assumptions for our long-term forecasts of AOP at years after the end of the chosen forecast horizon. For instance, if we believe that the firm can enlarge its abnormal profits beyond the short run, we may make a simple assumption that, after the end of the forecast horizon, AOP will grow at a constant rate (which we term terminal growth rate) in perpetuity. We then refine the residual operating income model as:

$$V_0^{NOA} = NOA_0 + \frac{AOP_1}{(1 + WACC)} + \frac{AOP_2}{(1 + WACC)^2} + \dots + \frac{AOP_T}{(1 + WACC)^T} + \frac{CV_T}{(1 + WACC)^T} \quad (15)$$

where V_0^{NOA} is the value of net operating assets, standing for the value of a firm to both shareholders and debtholders. The value of equity is then estimated via: $V_0^E = V_0^{NOA} - V_0^{NFO}$, where V_0^{NFO} is the value of net financial obligation. The subscript

T in Model (15) represents the number of years over the finite forecast horizon; CV_T is the continuing value beyond the chosen forecast horizon. Provided that AOP will grow permanently at a rate of g beyond the short-term forecast horizon, $CV_T = \frac{AOP_{T+1}}{WACC - g} = \frac{AOP_T * (1+g)}{WACC - g}$. On the other hand, if we assume AOP to be zero after the end of the forecast horizon, $CV_T = 0$; if we presume that AOP will remain constant and perpetual since the final year of the forecast horizon, $CV_T = \frac{AOP_{T+1}}{WACC} = \frac{AOP_T}{WACC}$. To facilitate more accurate forecasts of the terminal growth rate, we may further assume that, after the end of the finite forecast horizon, NOPM and ATO remain constant, and sales will grow at a constant rate or remain unchanged. Under such an assumption, AOP is a linear function of sales, thus the terminal growth rate in AOP equals the terminal growth rate in sales. As such, forecasts of the terminal growth in AOP are equivalent to forecasts of the terminal growth in sales, yet the latter will be more feasible for us.

After we apply the residual operating income model and estimate the equity value per share, we compare it with the stock price to aid in our investment decision-making. If the estimated equity value per share is significantly higher (lower) than the stock price, stock purchases (sales) are recommended. If the estimated equity value per share is insignificantly different from the stock price, it is advised to hold the stocks. Various risks which a firm is exposed to will make its future sales, ATO, NOPM, WACC, and terminal growth rate differ from the forecasted ones, increasing the uncertainty of, and error in, our forecasts and valuation. Thus, we need to perform sensitivity analysis in the meantime to see how sensitive the estimated equity value per share is to the changes in the forecast inputs across a certain range.⁷ Provided that the estimated equity value per share is highly sensitive to the forecast inputs and that we are uncertain about the accuracy of our forecasts, then a high upwards (downwards) sensitivity of the estimated equity value per share should lead us towards making buy (sell) decisions, *ceteris paribus*, while a high symmetric sensitivity should lead us towards holding the stocks, *ceteris paribus*. With all that said, the investment decisions made based on the foregoing analysis of fundamental values of stocks are premised on investors holding long-term investment horizons. If investors hold short-term investment horizons and seek to arbitrage on stock return anomalies, the fundamental value of stocks will be of less relevance for making profitable stock trades (He 2021).

88.8 How to Adjust for the Effects of Accruals-Based and Real Earnings Management in Equity Valuation?

Upwards accruals-based earnings management is just like borrowing future earnings for current use, thus earnings would be reversed and decline in the subsequent periods. Say, firms that delay writing-off assets in the current year would have to

⁷For example, if we speculate that the firm's liquidity and solvency positions will plausibly deteriorate in the near future, we need to first gauge the extent to which WACC will increase in consequence, and then estimate how large the estimated equity value per share will decrease as a result of the increase in WACC. On the other hand, if we are certain that the firm's liquidity and solvency positions will deteriorate, we should factor this into our estimation of discount rate (i.e., WACC) rather than in the sensitivity analysis.

write them off in the future years. In cases when we discern a firm's accruals manipulation in our valuation, we need to ensure that our forecast horizon is long enough (often at least 5 years and preferably 8–10 years) to allow accruals manipulation to play out by itself, even if the firm is in a mature industry with a well-established product line, stable prices of input products, robust production technologies, and stable market demand. Meanwhile, because accruals manipulation subjects a firm to a risk of detection and thereof to the related negative economic consequences, such risk needs to be factored into our estimation of the weighted average cost of capital (WACC) or into the test of the sensitivity of our estimated equity value per share to the WACC. Also, we need to adjust the weight we put on the current figures that are used in our forecasting.

Suppose we use the residual operating income model to value a firm's equity. If we spot that a firm has accelerated recognition of sales revenues in the current year, we should put less weight on the current sales revenues (current NOPM) that are used in the forecasts of future sales (future NOPM), and put more (less) weight on the current ATO that is used for forecasting future ATO if current ATO is higher (lower) than 1. If we detect an inflation of operating accruals via asset or liability distortion, we should put less (more) weight on the current NOPM (current ATO) when using it to forecast future NOPM (future ATO). Figure 1 illustrates how we should adjust the weight put on the current value-oriented figures when using them to forecast their future figures.⁸

Real earnings management subjects a firm to a relatively low risk of detection. As such, we don't need to adjust the discount rate (i.e., WACC) for the effects of real earnings management. However, unlike accruals-based earnings management, real earnings management that deviates a firm's optimal business strategies has negative real effects on a firm's future sales, NOPM, and ATO. Hence, our forecasts of future sales, NOPM, and ATO need to be adjusted downwards for the effects of the real earnings management; in the meantime, we need to put less weight on some of the current figures (i.e., sales and NOPM) used to forecast their future figures. Figure 2 portrays how to do these in the case when we use the residual operating income model to value a firm's equity.⁹

In specific, when we discern that managers have accelerated the timing of sales through giving increased price discounts or more lenient credit terms to customers, we should put less weight on the current sales that are used for forecasting future sales; then for the future sales forecasted based on current sales, we further adjust the future sales downwards. When we spot that a firm strategically reports lower costs of goods sold by virtue of overproduction, we should put less weight on the current NOPM that is used in the forecasts of future NOPM, and further, adjust the forecasted future NOPM downwards. When we discover the

⁸The same rationale and approach can be applied when we assess the role of downwards accruals manipulation in our forecasting and valuation.

⁹We may use the similar approach to assess the role the downwards real earnings management plays in the forecasting and valuation.

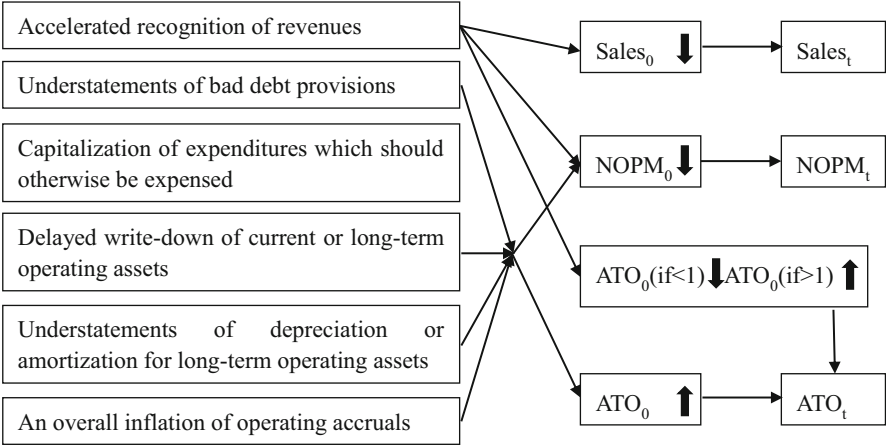


Fig. 1 Adjustment of accruals-based earnings management in equity valuation

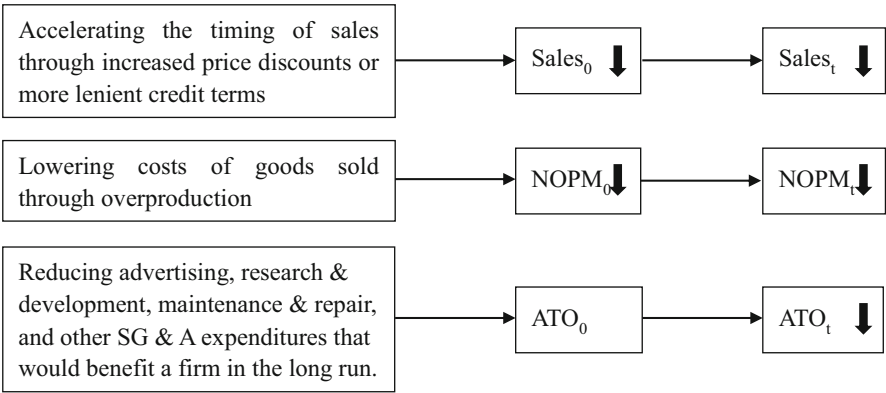


Fig. 2 Adjustment of real earnings management in earnings management

strategic reduction of advertising, research and development, maintenance and repair, and other selling, general, and administrative expenditures that would benefit a firm in the long run, we do not need to adjust the current ATO that is used for the forecasts of future ATO, but we need to adjust our forecasted future ATO downwards.

What if we see through real earnings management that is achieved via nonrecurring business activities (e.g., asset sales, debt-to-equity swaps, sales of securities that should have been aimed for long-term investments)? We don't need to make any adjustments in our forecasting, provided that the current value-oriented figures (e.g., ATO, NOPM) used for the forecasting are purged of nonrecurring operating items.

References

- Adam, T., and V.K. Goyal. 2008. The investment opportunity set and its proxy variables. *Journal of Financial Research* 31: 41–63.
- Altman, E.I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23: 589–609.
- Amiram, D., Z. Bozanic, J.D. Cox, Q. Dupont, J.M. Karpoff, and R. Sloan. 2018. Financial reporting fraud and other forms of misconduct: A multidisciplinary review of the literature. *Review of Accounting Studies* 23: 732–783.
- Armstrong, C.S., A.D. Jagolinzer, and D.F. Larcker. 2010. Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research* 48: 225–271.
- Armstrong, C.S., D.F. Larcker, G. Ormazabal, and D.J. Taylor. 2013. The relation between equity incentives and misreporting: The role of risk-taking incentives. *Journal of Financial Economics* 109: 327–350.
- Ashbaugh-Skaife, H., D.W. Collins, and W.R. Kinney. 2007. The discovery and reporting of internal control deficiencies prior to SOX-mandated audits. *Journal of Accounting and Economics* 44: 166–192.
- Badertscher, B.A. 2011. Overvaluation and the choice of alternative earnings management mechanisms. *The Accounting Review* 86: 1491–1518.
- Bagnoli, M., and S.G. Watts. 2010. Oligopoly, disclosure, and earnings management. *The Accounting Review* 85: 1191–1214.
- Baker, T., D. Collins, and A. Reitenga. 2003. Stock option compensation and earnings management incentives. *Journal of Accounting, Auditing and Finance* 18: 557–582.
- Balakrishnan, K. and D. A. Cohen. 2013. Competition and financial accounting misreporting. Available at SSRN 1927427.
- Ball, R., and L. Shivakumar. 2008. Earnings quality at initial public offerings. *Journal of Accounting and Economics* 45: 324–349.
- Balsam, S., I.-M. Haw, and S.B. Lilien. 1995. Mandated accounting changes and managerial discretion. *Journal of Accounting and Economics* 20: 3–29.
- Balsam, S., J. Krishnan, and J.S. Yang. 2003. Auditor industry specialization and earnings quality. *Auditing: A Journal of Practice & Theory* 22: 71–97.
- Bange, M.M., and W.F.M. De Bondt. 1998. R&D budgets and corporate earnings targets. *Journal of Corporate Finance* 4: 153–184.
- Barth, M.E., J.A. Elliott, and M.W. Finn. 1999. Market rewards associated with patterns of increasing earnings. *Journal of Accounting Research* 37: 387–413.
- Bartov, E. 1993. The timing of asset sales and earnings manipulation. *The Accounting Review* 68: 840–855.
- Bartov, E., and P. Mohanram. 2004. Private information, earnings manipulations, and executive stock-option exercises. *The Accounting Review* 79: 889–920.
- Beatty, A., and J. Weber. 2003. The effects of debt contracting on voluntary accounting method changes. *The Accounting Review* 78: 119–142.
- Becker, C.L., M.L. DeFond, J. Jambalvo, and K.R. Subramanyam. 1998. The effect of audit quality on earnings management. *Contemporary Accounting Research* 15: 1–24.
- Bedard, J.C., and K.M. Johnstone. 2004. Earnings manipulation risk, corporate governance risk, and auditors' planning and pricing decisions. *The Accounting Review* 79: 277–304.
- Beneish, M.D. 1999. The detection of earnings manipulation. *Financial Analysts Journal* 55: 24–36.
- Beneish, M.D., and M.E. Vargus. 2002. Insider trading, earnings quality, and accrual mispricing. *The Accounting Review* 77: 755–791.
- Beneish, M.D., E. Press, and M.E. Vargus. 2012. Insider trading and earnings management in distressed firms. *Contemporary Accounting Research* 29: 191–220.
- Bergstresser, D., and T. Philippon. 2006. CEO incentives and earnings management. *Journal of Financial Economics* 80: 511–529.

- Bharath, S.T., J. Sunder, and S.V. Sunder. 2008. Accounting quality and debt contracting. *The Accounting Review* 83: 1–28.
- Bhattacharya, N., E.L. Black, T.E. Christensen, and C.R. Larson. 2003. Assessing the relative informativeness and permanence of pro forma earnings and GAAP operating earnings. *Journal of Accounting and Economics* 36: 285–319.
- Bhattacharya, N., F. Ecker, P.M. Olsson, and K. Schipper. 2011. Direct and mediated associations among earnings quality, information asymmetry, and the cost of equity. *The Accounting Review* 87: 449–482.
- Bhojraj, S., P. Hribar, M. Picconi, and J. McInnis. 2009. Making sense of cents: An examination of firms that marginally miss or beat analyst forecasts. *The Journal of Finance* 64: 2361–2388.
- Biddle, G.C., G.S. Seow, and A.F. Siegel. 1995. Relative versus incremental information content *. *Contemporary Accounting Research* 12.
- Burgstahler, D., and I. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics* 24: 99–126.
- Burnett, B.M., B.M. Cripe, G.W. Martin, and B.P. McAllister. 2012. Audit quality and the trade-off between accretive stock repurchases and accrual-based earnings management. *The Accounting Review* 87: 1861–1884.
- Burns, N., and S. Kedia. 2006. The impact of performance-based compensation on misreporting. *Journal of Financial Economics* 79: 35–67.
- Bushee, B.J. 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73: 305–333.
- Campbell, J.Y., J. Hilscher, and J. Szilagyi. 2008. In search of distress risk. *The Journal of Finance* 63: 2899–2939.
- Campello, M., J.R. Graham, and C.R. Harvey. 2010. The real effects of financial constraints: Evidence from a financial crisis. *Journal of Financial Economics* 97: 470–487.
- Carcello, J.V., C.W. Hollingsworth, and T.L. Neal. 2006. Audit committee financial experts: A closer examination using firm designations. *Accounting Horizons* 20: 351–373.
- Chan, L.H., K.C.W. Chen, T.Y. Chen, and Y. Yu. 2014. Substitution between real and accruals-based earnings management after voluntary adoption of compensation Clawback provisions. *The Accounting Review* 90: 147–174.
- Cheng, Q., and D.B. Farber. 2008. Earnings restatements, changes in CEO compensation, and firm performance. *The Accounting Review* 83: 1217–1250.
- Cheng, Q., and T.D. Warfield. 2005. Equity incentives and earnings management. *The Accounting Review* 80: 441–476.
- Cheng, P., P. Man, and C.H. Yi. 2013. The impact of product market competition on earnings quality. *Accounting and Finance* 53: 137–162.
- Chi, W., L.L. Liscic, and M. Pevzner. 2011. Is enhanced audit quality associated with greater real earnings management? *Accounting Horizons* 25: 315–335.
- Chung, R., M. Firth, and J.-B. Kim. 2002. Institutional monitoring and opportunistic earnings management. *Journal of Corporate Finance* 8: 29–48.
- . 2005. Earnings management, surplus free cash flow, and external monitoring. *Journal of Business Research* 58: 766–776.
- Cohen, D.A., and P. Zarowin. 2010. Accrual-based and real earnings management activities around seasoned equity offerings. *Journal of Accounting and Economics* 50: 2–19.
- Cohen, D.A., A. Dey, and T.Z. Lys. 2008. Real and accrual-based earnings management in the pre- and post-Sarbanes-Oxley periods. *The Accounting Review* 83: 757–787.
- Coles, J.L., N.D. Daniel, and L. Naveen. 2006. Managerial incentives and risk-taking. *Journal of Financial Economics* 79: 431–468.
- Cornett, M.M., A.J. Marcus, and H. Tehranian. 2008. Corporate governance and pay-for-performance: The impact of earnings management. *Journal of Financial Economics* 87: 357–373.
- Damodaran, A. 2002. *Investment valuation*. New York: Wiley.
- Darrough, M., and S. Rangan. 2005. Do insiders manipulate earnings when they sell their shares in an initial public offering? *Journal of Accounting Research* 43: 1–33.

- DeAngelo, L.E. 1986. Accounting numbers as market valuation substitutes: A study of management buyouts of public stockholders. *The Accounting Review* 61: 400–420.
- Dechow, P.M., and I.D. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* 77: 35–59.
- Dechow, P.M., and C. Shakespear. 2009. Do managers time securitization transactions to obtain accounting benefits? *The Accounting Review* 84: 99–132.
- Dechow, P.M., R.G. Sloan, and A.P. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70: 193–225.
- Dechow, P.M., A.P. Hutton, and R.G. Sloan. 1996. Economic consequences of accounting for stock-based compensation. *Journal of Accounting Research* 34: 1–20.
- Dechow, P.M., S.P. Kothari, and R.L. Watts. 1998. The relation between earnings and cash flows. *Journal of Accounting and Economics* 25: 133–168.
- DeFond, M.L., and J. Jiambalvo. 1994. Debt covenant violation and manipulation of accruals. *Journal of Accounting and Economics* 17: 145–176.
- DeFond, M.L., and C.W. Park. 1997. Smoothing income in anticipation of future earnings. *Journal of Accounting and Economics* 23: 115–139.
- DeGeorge, F., Y. Ding, T. Jeanjean, and H. Stolowy. 2013. Analyst coverage, earnings management and financial development: An international study. *Journal of Accounting and Public Policy* 32: 1–25.
- Denis, D.J., and V. Sibikov. 2010. Financial constraints, investment, and the value of cash holdings. *Review of Financial Studies* 23: 247–269.
- Desai, M.A., A. Dyck, and L. Zingales. 2007. Theft and taxes. *Journal of Financial Economics* 84: 591–623.
- Dhaliwal, D., S. Huang, I.K. Khurana, and R. Pereira. 2014. Product market competition and conditional conservatism. *Review of Accounting Studies* 19: 1309–1345.
- Dichev, I.D., and D.J. Skinner. 2002. Large-sample evidence on the debt covenant hypothesis. *Journal of Accounting Research* 40: 1091–1123.
- Doyle, J.T., W. Ge, and S. McVay. 2007. Accruals quality and internal control over financial reporting. *The Accounting Review* 82: 1141–1170.
- Dyck, A., and L. Zingales. 2004. Private benefits of control: An international comparison. *The Journal of Finance* 59: 537–600.
- Dyck, A., A. Morse, and L. Zingales. 2010. Who blows the whistle on corporate fraud? *The Journal of Finance* 65: 2213–2253.
- Edwards, E.O., and P.W. Bell. 1961. *The theory and measurement of business income*. Berkeley/Los Angeles: University of California Press.
- Efendi, J., A. Srivastava, and E.P. Swanson. 2007. Why do corporate managers misstate financial statements? The role of option compensation and other factors. *Journal of Financial Economics* 85: 667–708.
- Erickson, M., and S.-w. Wang. 1999. Earnings management by acquiring firms in stock for stock mergers. *Journal of Accounting and Economics* 27: 149–176.
- Erickson, M., M. Hanlon, and E.L. Maydew. 2006. Is there a link between executive equity incentives and accounting fraud? *Journal of Accounting Research* 44: 113–143.
- Feng, M., W. Ge, S. Luo, and T. Shevlin. 2011. Why do CFOs become involved in material accounting manipulations? *Journal of Accounting and Economics* 51: 21–36.
- Francis, J.R., E.L. Maydew, and H.C. Sparks. 1999. The role of big 6 auditors in the credible reporting of accruals. *Auditing: A Journal of Practice & Theory* 18: 17–34.
- Francis, J., K. Schipper, and L. Vincent. 2003. The relative and incremental explanatory power of earnings and alternative (to earnings) performance measures for returns. *Contemporary Accounting Research* 20: 121–164.
- Francis, J., R. LaFond, P.M. Olsson, and K. Schipper. 2004. Costs of equity and earnings attributes. *The Accounting Review* 79: 967–1010.
- Francis, J., R. LaFond, P. Olsson, and K. Schipper. 2005. The market pricing of accruals quality. *Journal of Accounting and Economics* 39: 295–327.

- Frank, M.Z., and V.K. Goyal. 2009. Capital structure decisions: Which factors are reliably important? *Financial Management* 38: 1–37.
- Gande, A., and C.M. Lewis. 2009. Shareholder-initiated class action lawsuits: Shareholder wealth effects and industry spillovers. *The Journal of Financial and Quantitative Analysis* 44: 823–850.
- Ge, W., and S. McVay. 2005. The disclosure of material weaknesses in internal control after the Sarbanes-Oxley Act. *Accounting Horizons* 19: 137–158.
- Giroud, X., and H.M. Mueller. 2011. Corporate governance, product market competition, and equity prices. *The Journal of Finance* 66: 563–600.
- Gong, G., H. Louis, and A.X. Sun. 2008. Earnings management and firm performance following open-market repurchases. *The Journal of Finance* 63: 947–986.
- Gordon, M.J. 1962. *The investment, financing, and valuation of the corporation*. Homewood: R.D. Irwin.
- Graham, J.R., C.R. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40: 3–73.
- Guadalupe, M. and F. Pérez-González. 2010. Competition and private benefits of control. In *AFA 2007 Chicago Meetings Paper*.
- Gunny, K.A. 2010. The relation between earnings management using real activities manipulation and future performance: Evidence from meeting earnings benchmarks. *Contemporary Accounting Research* 27: 855–888.
- Hadlock, C.J., and J.R. Pierce. 2010. New evidence on measuring financial constraints: Moving beyond the KZ index. *The Review of Financial Studies* 23: 1909–1940.
- Hand, J.R.M. 1989. 1988 competitive manuscript award: Did firms undertake debt-equity swaps for an accounting paper profit or true financial gain? *The Accounting Review* 64: 587–623.
- Harris, M. 1998. The association between competition and managers' business segment reporting decisions. *Journal of Accounting Research* 36: 111–128.
- Haw, I.-M., B. Hu, L.-S. Hwang, and W. Wu. 2004. Ultimate ownership, income management, and legal and extra-legal institutions. *Journal of Accounting Research* 42: 423–462.
- He, G. 2015. The effect of CEO inside debt holdings on financial reporting quality. *Review of Accounting Studies* 20: 501–536.
- . 2016. Fiscal support and earnings management. *The International Journal of Accounting* 51: 57–84.
- . 2018. Credit ratings and managerial voluntary disclosures. *The Financial Review* 53: 337–378.
- . 2021. Credit rating, post-earnings-announcement drift, and arbitrage from transient institutions. *Journal of Business Finance & Accounting*. <https://doi.org/10.1111/jbfa.12520>.
- He, G. and H. Ren. 2021. *Are financially constrained firms more susceptible to a stock price crash?*. Working paper. <https://doi.org/10.2139/ssrn.3021307>.
- He, G., L. Bai, and H.M. Ren. 2019a. Analyst coverage and future stock Price crash risk. *Journal of Applied Accounting Research* 20: 63–77.
- He, G., D. Marginson, and X. Dai. 2019b. Do voluntary disclosures of product and business expansion plans impact analyst coverage and forecasts? *Accounting and Business Research* 49: 785–817.
- He, G., Z. A. Li and R. Slack. 2020a. *Does earnings management adversely affect analyst coverage and forecasts?*. Working paper. Durham University.
- He, G., H.M. Ren, and R. Taffler. 2020b. The impact of corporate tax avoidance on analyst coverage and forecasts. *Review of Quantitative Finance and Accounting* 54: 447–477.
- . 2021. Do corporate insiders trade on future stock price crash risk? *Review of Quantitative Finance and Accounting*. <https://doi.org/10.1007/s11156-020-00936-3>.
- Healy, P.M. 1985. The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics* 7: 85–107.
- Healy, P.M., and K.G. Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31: 405–440.

- Herrmann, D., T. Inoue, and W.B. Thomas. 2003. The sale of assets to manage earnings in Japan. *Journal of Accounting Research* 41: 89–108.
- Holthausen, R.W., D.F. Larcker, and R.G. Sloan. 1995. Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics* 19: 29–74.
- Hovakimian, A., T. Opler, and S. Titman. 2001. The debt-equity choice. *The Journal of Financial and Quantitative Analysis* 36: 1–24.
- Hribar, P., N.T. Jenkins, and W.B. Johnson. 2006. Stock repurchases as an earnings management device. *Journal of Accounting and Economics* 41: 3–27.
- Hunt, A., S. Moyer and T. Shevlin. 1997. Earnings volatility, earnings management and equity value. Working paper. University of Washington. <http://faculty.washington.edu/shevlin/hms2vol.pdf>.
- Jaggi, B., and P. Lee. 2002. Earnings management response to debt covenant violations and debt restructuring. *Journal of Accounting, Auditing and Finance* 17: 295–324.
- Jensen, M.C., and W.H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3: 305–360.
- Jiang, J., K.R. Petroni, and I. Yanyan Wang. 2010. CFOs and CEOs: Who have the most influence on earnings management? *Journal of Financial Economics* 96: 513–526.
- Jones, J.J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29: 193–228.
- Kaldor, N.C. 1966. Marginal productivity and the macro-economic theories of distribution: Comment on Samuelson and Modigliani. *The Review of Economic Studies* 33: 309–319.
- Kaplan, S.N., and L. Zingales. 1997. Do financing constraints explain why investment is correlated with cash flow? *Quarterly Journal of Economics* 112: 169–215.
- Karpoff, J.M., D.S. Lee, and G.S. Martin. 2008. The cost to firms of cooking the books. *Journal of Financial and Quantitative Analysis* 43: 581–611.
- Karuna, C. 2007. Industry product market competition and managerial incentives. *Journal of Accounting and Economics* 43: 275–297.
- Karuna, C., K. Subramanyam and F. Tian. 2015. *Competition and earnings management*. Working paper. University of Southern California. <https://www.semanticscholar.org/paper/Competition-and-Earnings-Management-Karuna-Subramanyam/ad15f09d4909a9debf61abd31de2403b90a8868f>.
- Kasznik, R., and M.F. McNichols. 2002. Does meeting earnings expectations matter? Evidence from analyst forecast revisions and share prices. *Journal of Accounting Research* 40: 727–759.
- Kayhan, A., and S. Titman. 2007. Firms' histories and their capital structures. *Journal of Financial Economics* 83: 1–32.
- Ke, B., S. Huddart, and K. Petroni. 2003. What insiders know about future earnings and how they use it: Evidence from insider trades. *Journal of Accounting and Economics* 35: 315–346.
- Keating, A.S., and J.L. Zimmerman. 1999. Depreciation-policy changes: Tax, earnings management, and investment opportunity incentives. *Journal of Accounting and Economics* 28: 359–389.
- Kedia, S., and T. Philippon. 2009. The economics of fraudulent accounting. *The Review of Financial Studies* 22: 2169–2199.
- Kinney, W.R., and L.S. McDaniel. 1989. Characteristics of firms correcting previously reported quarterly earnings. *Journal of Accounting and Economics* 11: 71–93.
- Kohlbeck, M., and X. Luo. 2019. Are CFO debt-like compensation incentives associated with financial reporting quality? *Advances in Accounting* 45: 100413.
- Korajczyk, R.A., and A. Levy. 2003. Capital structure choice: Macroeconomic conditions and financial constraints. *Journal of Financial Economics* 68: 75–109.
- Kothari, S.P., A.J. Leone, and C.E. Wasley. 2005. Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39(1): 163–197. <https://doi.org/10.1016/j.jacceco.2004.11.002>.
- Kothari, S.P., N. Mizik, and S. Roychowdhury. 2016. Managing for the moment: The role of earnings management via real activities versus accruals in SEO valuation. *The Accounting Review* 91: 559–586.

- Krishnan, G.V. 2003. Does big 6 auditor industry expertise constrain earnings management? *Accounting Horizons* 17: 1–16.
- Kumar, P., and N. Langberg. 2009. Corporate fraud and investment distortions in efficient capital markets. *The Rand Journal of Economics* 40: 144–172.
- Larcker, D.F., S.A. Richardson, and I.R. Tuna. 2007. Corporate governance, accounting outcomes, and organizational performance. *The Accounting Review* 82: 963–1008.
- Lee, C.M.C., J. Myers, and B. Swaminathan. 1999. What is the intrinsic value of the Dow. *Journal of Finance* 54: 1693–1741.
- Leland, H.E. 1998. Agency costs, risk management, and capital structure. *The Journal of Finance* 53: 1213–1243.
- Leuz, C., D. Nanda, and P.D. Wysocki. 2003. Earnings management and investor protection: An international comparison. *Journal of Financial Economics* 69: 505–527.
- Linck, J.S., J. Netter, and T. Shu. 2013. Can managers use discretionary accruals to ease financial constraints? Evidence from discretionary accruals prior to investment. *The Accounting Review* 88: 2117–2143.
- Loebbecke, J.K., M.M. Eining, and J.J. Willingham. 1989. Auditors experience with material irregularities-frequency, nature, and detectability. *Auditing: A Journal of Practice & Theory* 9: 1–28.
- Louis, H. 2004. Earnings management and the market performance of acquiring firms. *Journal of Financial Economics* 74: 121–148.
- Low, A. 2009. Managerial risk-taking behavior and equity-based compensation. *Journal of Financial Economics* 92: 470–490.
- Maksimovic, V., and S. Titman. 1991. Financial policy and reputation for product quality. *The Review of Financial Studies* 4: 175–200.
- Markarian, G., and J. Santalo'. 2014. Product market competition, information and earnings management. *Journal of Business Finance & Accounting* 41: 572–599.
- Mayers, S. and N. Majluf. 1984. Corporate financing and investment decision, when firms have decisions that investors do not have. *Journal of Financial Economics* 13(2): 187–221.
- McNichols, M.F., and S.R. Stubben. 2008. Does earnings management affect firms' investment decisions? *The Accounting Review* 83: 1571–1603.
- Miller, G.S. 2006. The press as a watchdog for accounting fraud. *Journal of Accounting Research* 44: 1001–1033.
- Mitra, S., and W.M. Cready. 2005. Institutional stock ownership, accrual management, and information environment. *Journal of Accounting, Auditing and Finance* 20: 257–286.
- Modigliani, F., and M.H. Miller. 1958. The cost of capital, corporation finance and the theory of investment. *The American Economic Review* 48: 261–297.
- . 1963. Corporate income taxes and the cost of capital: A correction. *The American Economic Review* 53: 433–443.
- Myers, J.N., L.A. Myers, and D.J. Skinner. 2007. Earnings momentum and earnings management. *Journal of Accounting, Auditing and Finance* 22: 249–284.
- Nenova, T. 2003. The value of corporate voting rights and control: A cross-country analysis. *Journal of Financial Economics* 68: 325–351.
- Nissim, D., and S.H. Penman. 2001. Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies* 6: 109–154.
- Ohlson, J.A. 1995. Earnings, book values and dividends in equity valuation. *Contemporary Accounting Research* 11: 661–687.
- Park, M.S., and T. Park. 2004. Insider sales and earnings management. *Journal of Accounting and Public Policy* 23: 381–411.
- Peasnell, K.V. 1982. Some formal connections between economic values and yields and accounting numbers. *Journal of Business Finance and Accounting* 9: 361–381.
- Penman, S.H., and X.J. Zhang. 2002. Accounting conservatism, the quality of earnings, and stock returns. *The Accounting Review* 77: 237–264.
- Petroni, K.R. 1992. Optimistic reporting in the property- casualty insurance industry. *Journal of Accounting and Economics* 15: 485–508.

- Qi, C., K. Subramanyam and J. Zhang. 2010. *Accrual quality, bond liquidity, and cost of debt*. Working paper. University of Southern California. <https://www.marshall.usc.edu/sites/default/files/jieying/intellcont/QSZ20100827-1.pdf>.
- Rangan, S. 1998. Earnings management and the performance of seasoned equity offerings. *Journal of Financial Economics* 50: 101–122.
- Richardson, S.A., R.G. Sloan, M.T. Soliman, and İ. Tuna. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics* 39: 437–485.
- Robb, S.W.G. 1998. The effect of analysts' forecasts on earnings management in financial institutions. *Journal of Financial Research* 21: 315–331.
- Rosner, R.L. 2003. Earnings manipulation in failing firms. *Contemporary Accounting Research* 20: 361–408.
- Rountree, B., J.P. Weston, and G. Allayannis. 2008. Do investors value smooth performance? *Journal of Financial Economics* 90: 237–251.
- Roychowdhury, S. 2006. Earnings management through real activities manipulation. *Journal of Accounting and Economics* 42: 335–370.
- Sadka, R. 2006. Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. *Journal of Financial Economics* 80: 309–349.
- Saleh, N.M., and K. Ahmed. 2005. Earnings management of distressed firms during debt renegotiation. *Accounting and Business Research* 35: 69–86.
- Sawicki, J., and K. Shrestha. 2014. Misvaluation and insider trading incentives for accrual-based and real earnings management. *Journal of Business Finance & Accounting* 41: 926–949.
- Schilit, H. 2010. *Financial shenanigans*. Maidenhead: Tata McGraw-Hill Education.
- Shi, G., J. Sun, and L. Zhang. 2018. Product market competition and earnings management: A firm-level analysis. *Journal of Business Finance & Accounting* 45: 604–624.
- Shivakumar, L. 2000. Do firms mislead investors by overstating earnings before seasoned equity offerings? *Journal of Accounting and Economics* 29: 339–371.
- Shleifer, A. 2004. Does competition destroy ethical behavior? *American Economic Review* 94: 414–418.
- Sickles, R.C., and V. Zelenyuk. 2019. *Measurement of productivity and efficiency: Theory and practice*. Cambridge: Cambridge University Press.
- Stein, J.C. 1989. Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *The Quarterly Journal of Economics* 104: 655–669.
- Sweeney, A.P. 1994. Debt-covenant violations and managers' accounting responses. *Journal of Accounting and Economics* 17: 281–308.
- Teoh, S.H., I. Welch, and T.J. Wong. 1998. Earnings management and the long-run market performance of initial public offerings. *The Journal of Finance* 53: 1935–1974.
- Whited, T.M., and G. Wu. 2006. Financial constraints risk. *The Review of Financial Studies* 19: 531–559.
- Yu, F. 2008. Analyst coverage and earnings management. *Journal of Financial Economics* 88(2): 245–271. <https://doi.org/10.1016/j.jfineco.2007.05.008>.
- Zang, A.Y. 2012. Evidence on the trade-off between real activities manipulation and accrual-based earnings management. *The Accounting Review* 87: 675–703.



The Applications of Machine Learning in Accounting and Auditing Research

89

Hanxin Hu and Ting Sun

Abstract

The term “machine learning” has become a buzzword in the past few years. In accounting and auditing area, while this technology has been used in major accounting firms such as Big 4 s, its research is still evolving. Increased use of machine learning and other artificial intelligence techniques will allow accountants to focus on providing better decision support instead of on data gathering and manual analyses. This chapter introduces machine learning as compared to traditional statistical modeling, discusses its current applications in accounting and auditing research, and provides directions for future research.

Keywords

Machine learning · Artificial intelligence · Accounting · Auditing · Data analytics

JEL Classification

M41 · M42

89.1 Machine Learning Technology

Machine learning is an application of artificial intelligence (AI) that enables computer systems to automatically learn patterns and regularities from historical data and improve from past experience without human intervention or adjust actions

H. Hu

Department of Accounting and Information Systems, Rutgers University, Newark, NJ, USA
e-mail: hh497@scarletmail.rutgers.edu

T. Sun (✉)

Department of Accounting and Information Systems, The College of New Jersey, Ewing, NJ, USA
e-mail: sunt@tcnj.edu

(Alpaydin 2020; Expert.ai 2020). The identified patterns and regularities can help one make predictions for future unseen data, assuming that the future will not be much different from the past when the data was collected (Alpaydin 2020).

A machine learning model is created directly from data by an algorithm. It is often called by researchers “black box” because it cannot explain how input variables are being combined to make predictions in a way that humans can understand. Most machine learning models, however, are not designed with interpretability constraints. Instead, their goal is to make accurate predictions. In his book “Introduction to Machine Learning,” Ethem Alpaydin (2020) states that:

“We may not be able to identify the process completely, but we believe we can construct a good and useful approximation. That approximation may not explain everything, but may still be able to account for some part of the data. We believe that though identifying the complete process may not be possible, we can still detect certain patterns or regularities. This is the niche of machine learning.”

In 1949, Donald Hebb, in his book titled “The Organization of Behavior” (Hebb 1949), developed a model of brain cell interaction, which established a foundation for the development of machine learning. The AI pioneer, Arthur Samuel, first came up with the term “Machine Learning” in 1952 and popularized it. He invented the first computer program for playing checkers at IBM. The program was among the world’s first successful self-learning program (Wiederhold et al. 1990). Using the positions of the pieces on the checkerboard, he designed a scoring function that is capable of measuring the winning chances for each side. The checkers-playing program decides its next move by employing a minimax strategy, which eventually evolved into the minimax algorithm. It records all positions that have been shown and combines each position with the values of the reward function (Foote 2019). In 1957, Frank Rosenblatt related Donald Hebb’s model of brain cell interaction to Arthur Samuel’s machine learning concept and created the Perceptron, which was a simple classifier for image recognition in an artificial neural network. It was initially designed for IBM 704, a 5-ton computer with the size of a room. With perceptions, IBM704 can learn by itself to distinguish punch cards marked on the left from those marked on the right. Rosenblatt (1957) called it “a machine capable of perceiving, recognizing and identifying its surroundings without any human training or control (Lefkowitz 2019).” Due to the limitation of hardware and technology at that time, perceptions could not recognize many kinds of visual patterns such as human faces, then there were several years of stagnation of the neural network area until a resurgence during the 1990s (Foote 2019).

Machine learning regained its popularity because the intersection of computer science and statistics generated probabilistic approaches in AI in the 1990s. The basic idea behind it is that learning can be considered as inferring plausible models for explaining observed data and making predictions. Nevertheless, there is an uncertainty regarding which model is appropriate, given the observed data is uncertain. Likewise, predictions about future unseen data are uncertain. Probability theory provides a framework to represent and manipulate uncertainty about models and predictions (Ghahramani 2015), and, thereby, emerges as one of the principal approaches for machine learning, shifting the focus of this technology toward data-driven approaches (Provalis Research 2017).

With large-scale data available as well as the advances in computing and the rapid growth of data storage techniques, machine learning enters its golden eras in recent decades. Machine learning algorithms such as deep learning have been evolved to be capable of analyzing big data. As a highlight, DeepMind’s AlphaGo defeated a world champion of Go game, which is known as the most challenging classical game for AI because of its complexity.

89.2 Differences Between Machine Learning and Statistical Modeling

Machine learning is built upon statistics. While closely related to each other, machine learning and statistical modelling are distinct in many aspects. Table 1 summarizes major differences between them. The biggest distinction is the principal goal: machine learning models are developed to make accurate predictions whereas statistical modeling is designed for inference about the relationships between input variables (Healthcare.ai 2020). In other words, statistical modeling draws population inferences from a data sample, whereas machine learning identifies generalizable predictive patterns (Bzdok et al. 2018). While statistical modeling does make predictions, the accuracy of those predictions made by statistical modelling is lower than those generated by machine learning approaches. Similarly, certain machine learning models like Lasso regression provide high level of interpretability; nevertheless, others like support vector machine are very difficult to explain.

In contrast to statistical modeling such as linear regression, machine learning approach does not require the data to follow any prior assumptions about the underlying relationships between the variables, such as normality of residuals, homoscedasticity, linearity of residuals, no multicollinearity, etc. This feature enables machine learning to make more accurate predictions because data in the real word does not always follow those rigid assumptions.

Machine learning performs excellently with high-dimensional data sets. The bigger the data size, the higher the accuracy of predictions. For example, deep neural networks typically cannot make good predictions when the sample size is limited, but the prediction accuracy improves with the increase of sample size. In contrast, statistical models generally predict accurately on hundreds or even dozens of data points but improve little with the addition of more observations (Healthcare.ai).

In traditional statistics, we typically do not want to include too many variables to avoid issues like multicollinearity. We use techniques such as estimates of the variance inflation factor (VIF) to identify the multicollinearity. However, the

Table 1 Machine learning versus statistical modeling

	Machine learning	Statistical modeling
Purpose	Prediction	Interpretation
Assumptions	No	Yes
Data	Big data	Small data
Variables	Many	Few

predictive power of a machine learning model will not be affected by the existence of highly colinear variables. Furthermore, machine learning algorithms have mechanisms to rule out irrelevant information. For instance, support vector machine allows more predictors than features because the kernel trick used by the algorithm allows the system to operate solely on the inner product of those feature vectors.

89.3 Machine Learning Algorithms

Machine learning algorithms are considered the engines of machine learning due to automatically transforming a dataset into a model (Heller 2019). Implementing machine learning algorithms can efficiently extract information and detect significant patterns underlying the data (Shalev-Shwartz and Ben-David 2014). As illustrated in Fig. 1, five categories of learning algorithms exist, including supervised learning, unsupervised learning, ensemble learning, neural networks and deep learning, and reinforcement learning.

The running of supervised learning algorithms can be expressed as a mapping function $y(x)$ that takes a new observation x as the input and generates a predicted output y (Bishop 2006). In a supervised learning paradigm, the dataset is divided into three compartments: the learning (training) set, validation set, and test set, separately for developing machine models, tuning the parameters in models, and examining the models' generalization effects. According to the nature of outputs, supervised learning algorithms can be further partitioned into two types: regression and classification. The outputs are real values within the regression scenario, where the algorithms, such as linear and polynomial regressions, are commonly applied. For instance, the New York City Airbnb listing prices can be predicted by using regression algorithms.¹ Unlike the regression, the dependent variables of classification models (e.g., logistic regression, support vector machine, and Naïve Bayes classifier) are in the form of categories. Consider the MNIST database of handwritten digits² as a classification example, where the input variables are pixel values, and the output is the identity of digits.

By contrast with supervised learning, no outputs correspond to input variables under the circumstance of unsupervised learning (Shaw 2017). That is, it is employed to explore the potential patterns hidden in the data without prior labels. There are three types of unsupervised learning algorithms: clustering, association rule learning, and dimensionality reduction. The clustering techniques' insight is to group a part of data samples based upon the essential principle: maximizing the intraclass similarity and minimizing the interclass similarity (Chen et al. 1996). It can be used in the marketing campaign by categorizing customers with similar

¹The data source is <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>. In this application case, explanatory variables including room types, geographical availability, and the number of reviews per month are treated as the independent variables to predict prices.

²The data source is <http://yann.lecun.com/exdb/mnist/>

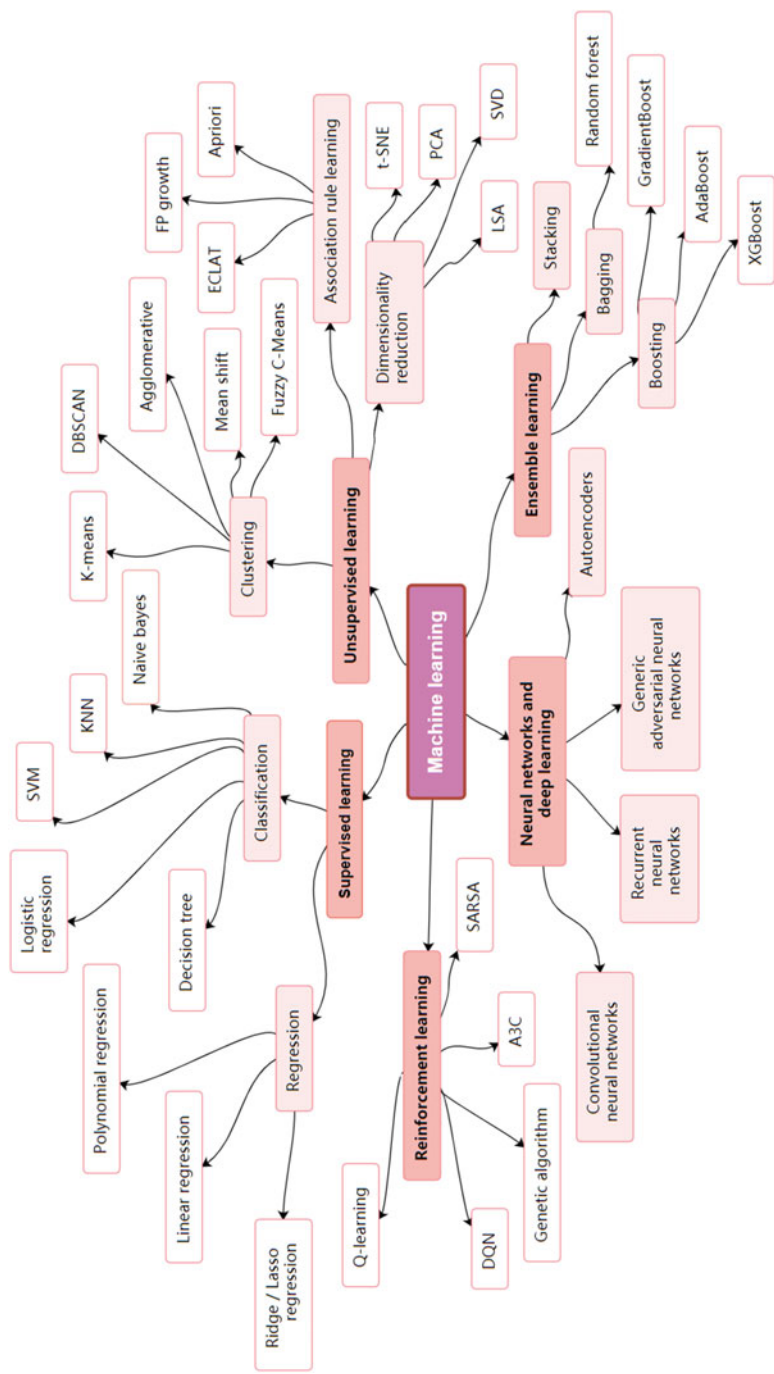


Fig. 1 The roadmap of machine learning algorithms

consumption behaviors and establishing appropriate segmentation strategies.³ Regarding the association rule learning that was initially developed by Agrawal et al. (1993), it attempts to capture meaningful associations and causal structure among items in the database (Zhao and Bhowmick 2003). One application for this kind of algorithms (e.g., Apriori association rule learning algorithm) is the construction of recommendation systems. For example, a customer has ordered diapers from the Amazon website; perhaps next time, powdered milk will be exhibited in his or her recommendation list due to the meaningful relationship between those two items mined by the association rule learning. As implied by the name, dimensionality reduction algorithms aim to reduce the number of predictors in the dataset but remain critical information in the meantime. This objective can be achieved either through feature selection or extraction (Khalid et al. 2014). Specifically, feature selection refers to the process where redundant and noisy attributes are filtered out, and the best subset of predictors are selected to represent the original high-dimensional dataset (e.g., backward or forward stepwise regression); while feature extraction, such as principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), involves several linear or nonlinear transformations of the original features to generate new attributes that contribute to latter learning tasks (i.e., classification or regression) (Khalid et al. (2014) and Van Der Maaten et al. (2009)).

The nature of the learning algorithms described above is searching a hypothesis space (i.e., training set) to find the best single learner (Dietterich 2002). Instead of only using one unique learner, ensemble learning combines the approximation results generated from multiple base learners (e.g., decision tree) (Zhou 2009). One of the effective and representative ensemble learning methods is bagging, in which each base learner is trained from a different bootstrap replicate of the learning set,⁴ and the final output aggregates their prediction results when predicting the numerical outcome and voting for the majority class (Breiman 1996). Relative to the single learner, ensemble learning usually achieves lower-generation errors on the previously unseen dataset, because they can partially overcome three problems afflicting the unique learning algorithm: the statistical problem,⁵ the computational problem,⁶ and the representation problem⁷ (Dietterich 2002).

³One of the examples in this domain is in <https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python>. The task of this data set is to segment customers based on their behavioral related attributes by applying the K-means clustering algorithms.

⁴A bootstrap replicate is accessed by randomly sampling the training set with replacement. This operation will generate a new training set with size equal to that of the original one.

⁵The statistical problem means that the training set fails to provide adequate information to select one single learner within the circumstance that multiple unique learners can achieve the same accuracy on the training set.

⁶Searching for the best hypothesis (e.g., neural network) that fits in the training data may be computationally intractable.

⁷The approximations to the real target function, which are generated from single learners, may not be ideal.

In the recent two decades, neural networks and deep learning are increasingly popular for applications ranging from stock price prediction to face recognition (Anthony and Bartlett 2009). These methods are considered the branch of machine learning algorithms: obtaining a nonlinear approximation to the true target function by learning successive layers of significant representations (Chollet 2017). Compared with traditional supervised learning algorithms, neural networks and deep learning can more accurately capture intricate patterns underlying the large datasets with the implementation of back-propagation algorithms to update the internal parameters contained in every layer (Lecun et al. 2015). Particularly, convolutional neural networks dramatically contribute to speech recognition and visual object recognition; recurrent neural networks are universally used in the field of text processing (Lecun et al. 2015).

Finally, reinforcement learning attempts to “maximize a numerical reward signal by mapping situations into actions” (Sutton and Barto 2018). In other words, this subfield of machine learning algorithms allows the agent to determine the best subsequent strategies given the current status by learning interactive behaviors that maximize the reward (Shaw 2017). Dissimilar with supervised and unsupervised learning, these algorithms’ goal is neither to obtain the optimal data representation nor find the hidden structure on unlabeled datasets (Bishop 2006; Sutton and Barto 2018). In practice, they can be applied in contexts like autonomous driving and robotics (Sallab et al. 2017; Kober et al. 2013).

89.4 Applications of Machine Learning in Accounting Research

Machine learning has been increasingly integrated into empirical accounting research to gauge information content of accounting amounts and qualitative disclosures, evaluate the corporations’ bankruptcy or default risks, and enhance accounting estimates & fundamental analytics (Bertomeu 2020).

89.4.1 The Evaluation of Information Content

Machine learning algorithms excel at adopting a flexible approach to efficiently extract meaningful information from various data formats, including numerical attributes and raw texts. This intrinsic nature sheds light on the evaluation of information content. Barth et al. (2021) use the nonlinear learning algorithm CART (i.e., classification and regression tree) to examine the relationship between equity prices and accounting amounts (e.g., net income, equity book value, research and development expenses, etc.). They discover that CART’s out-of-sample performance significantly outperforms that for the traditional linear regression model, suggesting that accounting amounts tend to map into future cash flows and equity prices nonlinearly. By comparing the model’s explanatory power over calendar years, this study contradicts the conclusion generated from prior research that the relevance between accounting information and equity prices gradually decays with the new economy’s development in the recent five decades.

Instead of focusing on quantitative accounting information (e.g., accounting amounts shown in financial restatements), some researchers derive insights from qualitative disclosures, such as the company's annual reports, earning call conference transcripts, and analyst reports. However, manually analyzing and classifying substantial amount of raw textual data are incredibly strenuous (Van Den Bogaerd and Aerts 2011). In this sense, machine learning algorithms are efficient and convenient in extracting meaningful information from voluminous texts. Li (2010) construct a Naïve Bayesian classifier to identify the tone of forward-looking statements (FLS) in the Management Discussion and Analysis section (MD&A) of 10-K and 10-Q reports. Even though the conditional dependence assumption behind this learning algorithm does not align with reality, the prediction accuracy yielded from cross validation is 67%, saliently better than random guessing. This chapter demonstrates the informativeness of MD&A by discovering the positive association between the average tone of the FLS and future earnings performance. Note that the author also uses dictionaries to obtain the tone but fails to forecast the company's future performance, indicating that machine learning algorithms may process corporate filings better than the traditional dictionary approach. Analogous to Li (2010), Huang et al. (2014) also exploit Naïve Bayesian to predict the opinion (i.e., positive, negative, and neutral) at the sentence level, yet based on another corpus analyst reports. This chapter demonstrates that after controlling the effects of quantitative items derived from analyst reports (e.g., earnings forecasts and stock recommendations), investors react significantly to qualitative information contained in this corpus. Concentrating on the qualitative disclosures of MD&A, Frankel et al. (2016) build a support vector regression (SVR) model to predict firm-level accruals given the occurrence frequency of the entire one- and two-word phrases used as independent variables. After incorporating this numerical output into the accrual model (Dechow and Dichev 2002), they find that the explanatory power of the modified model augments by around 24.8%, reflecting that the information content in MD&A interprets accruals in a way distinguished from that of Dechow and Dichev (2002) model.

Li (2010), Huang et al. (2014), and Frankel et al. (2016) all employ supervised learning techniques to obtain the real functions' approximations and predict the target labels or numbers. Nevertheless, manually labeling outcomes in the training set is time-consuming and vulnerable to programmers' subjective judgments (e.g., the taxonomy of risk factors disclosed on 10-K reports varies from person to person) (Bao and Datta 2014). Aiming at mitigating these issues, some other studies apply unsupervised learning algorithms to unravel the underlying structure of textual data. Huang et al. (2018) employ a Bayesian-parametric topic model latent Dirichlet allocation (LDA) to compare the thematic contents of conference call transcripts with those discussed in analyst reports instantly released after calls. Two critical findings of this study establish both the discovery and interpretation roles played by analyst reports during the information intermediary process: (1) investors significantly and economically react to the new and exclusive topics covered in analyst reports, especially when managers are confronted with greater incentives to conceal value-relevant information (e.g., undergoing undesirable financial performance) in earning call disclosures; (2) within the contexts where earning calls are demanding in comprehensibility, the explanations offered in analyst reports are particularly valuable to investors. With the contextual

assumption that words appearing in one sentence only refer to one risk factor, Bao and Datta (2014) modify the original topic model and implement a novel sent-LDA model to automatically capture risk categories embraced in section 1A of 10-K reports. This study then investigates whether and how risk perceptions proxied by stock volatilities respond to these risk-related disclosures. It is intriguing to observe that two-thirds of risk types exert insignificant influence (e.g., competition risks and volatile stock price risks). Only three informative risks related to systematic and liquidity risks tend to increase the risk perceptions of investors.

89.4.2 The Estimation of Bankruptcy or Default Risks

For credit risk management, there has been considerable research attempting to estimate the likelihood of bankruptcy or default events. Ohlson (1980) is one of the most seminal literature by applying the machine learning algorithm logistic regression to predict potential debtors' default probability. Afterward, a variety of more advanced learning methods are exploited in this domain to alleviate some unpreferable nature of logistic regression, such as (1) the incapability of capturing the nonlinear association between the target outcomes and predictors, (2) and the ineffectiveness when working on high-dimensional datasets. Table 2 summarizes representative bankruptcy or default prediction papers published in the recent three decades from the angles of dataset, dependent variables, predictors, machine learning algorithms, and out-of-sample performance. It is manifest that a wealth of studies emphasizes the improvement of model fitting by applying a wide range of machine learning algorithms (e.g., artificial neural networks, support vector machine, and ensemble learning methods) to achieve better prediction accuracy (Odom and Sharda (1990), Yang et al. (1999), Shin et al. (2005), Kim and Sohn (2010), Tsai et al. (2014), and Barboza et al. (2017)).

Nevertheless, this set of models is considered a "black box" due to the lack of transparency (Adadi and Berrada 2018). To mitigate this issue, Olson et al. (2012) employ the decision tree algorithm to generate interpretative rules during the bankruptcy prediction process. After controlling the decision tree algorithm's complexity,⁸ it can achieve a higher classification accuracy relative to support vector machine and artificial neural networks models.

It is worth noting that most literature shown in Table 2 solely uses quantitative financial items as the inputs of machine learning models. However, certain qualitative disclosures are also predictive of potential bankruptcy. Having complemented bankruptcy-relevant financial ratios with the token count vectors derived from MD&A sections, Cecchini et al. (2010a) find the prediction accuracy saliently improves.

⁸"Complexity" means that the decision tree model generates a plethora of rules, resulting the overfitting issues.

Table 2 Summary of the papers

Paper	Dataset	Dependent variables	Independent variables	Algorithms	Out-of-sample performance
Odom and Sharda (1990)	Moody's Industrial Manuals	Whether the firm went bankrupt between 1975 and 1982	Five financial ratios used by Altman (1968)	Artificial neural networks with three layers	ANN can achieve a consistently high true positive rate ^a with varying class ratio.
Yang et al. (1999)	US oil and gas industry data	Whether the firm went bankrupt between 1984 and 1989	Five financial ratios used by Platt et al. (1994)	Probabilistic neural networks with four layers	Neural networks have the best performance concerning the overall classification accuracy, but not in Type II errors ^b .
Shin et al. (2005)	Provided by Korea Credit Guarantee Fund	Whether the firm went bankrupt or not between 1996 and 1999	Ten financial ratios including total asset growth and contribution margin	Support vector machine and back-propagation neural networks with three layers	As the size of the training set gets smaller, SVM has higher classification accuracy than neural networks.
Kim and Sohn (2010)	Korean dataset	Whether small and medium enterprises defaulted or not between 1997 and 2002	Eight attributes of small and medium enterprises; 9 financial ratios; 16 technology evaluation factors; 10 economic indicators	Support vector machine, back-propagation neural networks with three layers, and logistic regression	SVM has the highest classification accuracy among the three learning algorithms.
Olson et al. (2012)	Compustat, LexisNexis databases	Whether the firm went bankrupt between January 2006 and December 2009	Sixteen financial items derived from prior research	Decision tree, support vector machine, back-propagation neural networks with three layers, and logistic regression	After controlling the complexity of the decision tree algorithm, it can achieve the highest classification accuracy.

(continued)

Table 2 (continued)

Paper	Dataset	Dependent variables	Independent variables	Algorithms	Out-of-sample performance
Tsai et al. (2014)	Taiwan Economic Journal	Whether the firm went bankrupt	Ninety-five relevant financial ratios	Bagging and boosting ensemble learning with different base classifiers	Decision tree ensemble by boosting outperforms other models according to prediction accuracy.
Barboza et al. (2017)	Compustat, NYU's Salomon Center databases	Whether the firm filed for bankruptcy between 1985 and 2014	Eleven quantitative financial items used by Altman (1968) and Carton and Hofer (2006)	Logistic regression, artificial neural networks, support vector machine with different kernels, boosting, bagging, and random forest	Random forest generates the highest AUC ROC score ^c compared with other algorithms.
Cecchini et al. (2010a)	Edgar, Compustat/CRSP databases	Whether the firm filed for bankruptcy between 1994 and 1999	The token count vectors generated from MD&A sections, financial items used by Altman (1968) and Beneish (1999)	Support vector machine	After merging the quantitative items with textual information, the prediction accuracy of bankruptcy improves.

^aTrue positive rate is defined as the likelihood that the actual positive instances (i.e., bankruptcy cases in this chapter) will test positive.

^bType II errors produce false negatives, where the bankrupt firms are wrongly classified into the non-bankruptcy group.

^cThe area under the curve (AUC) measures the performance of classification models at various cut-off settings.

89.4.3 The Enhancement of Accounting Estimates and Fundamental Analytics

By exerting the predictive role, machine learning algorithms can support judgments and decision-making processes (Cho et al. 2020). Specifically, they inherently fit in the contexts of accounting estimates and fundamental analytics, in which the likelihood of manager manipulation or subjective judgments is non-trivial. Grounded upon the data obtained from the insurance industry, Ding et al. (2020) apply four kinds of machine learning algorithms (i.e., linear regression, random forest, gradient boosting regression, and artificial neural networks) to

predict the amount of loss reserves. This study shows that compared with managers' subjective estimations, the approximations generated from machine learning algorithms are closer to real loss reserve values. To alleviate the problem that the out-of-sample performance of traditional regression models in predicting profitability does not significantly outperform that generated from random walking, Anand et al. (2019) use the random forest learning algorithm to forecast the directional changes of five profitability proxies (i.e., return on equity, return on assets, return on net operating assets, cash flow from operations, and free cash flow). By incorporating a minimum number of independent variables into the classification model (i.e., the firm identifier, fiscal year, and the contemporaneous value of the target profitability measure), this study substantiates that machine learning algorithms can achieve better prediction accuracy than the random walking benchmark over forecast windows ranging from 1 to 5 years. This finding is meaningful to the improvement of fundamental analytics, where the profitability prediction is one of the key procedures.

89.5 Applications of Machine Learning in Auditing Research

In the context of auditing, researchers have explored various machine learning's applications in audit processes, detecting financial misstatements, predicting audit quality, and etc.

89.5.1 Auditing Processes

There are many possible applications of machine learning in audit procedure from planning to reporting. Using automation for large volumes of data both structured and unstructured can help delivering high-quality audits, which allows auditors to focus more on tasks that require higher level of human judgment. For example, advanced machine learning techniques such as deep learning can analyze unstructured data such as audit working papers to identify potential errors and risks. Sun (2019) focuses on deep learning technique and proposes a framework for applying it to all audit phases. The proposed framework identifies that deep learning has the capabilities of text understanding, speech recognition, visual recognition, and structured data analysis, which can be fit into the audit environment. It also illustrates that audit processes like substantive testing can be automated by deep learning through its functions of information identification and judgment support. Examples include examining business contracts, emails, recordings, or minutes of board meetings to collect audit evidence. The implementation of machine learning is leading the accounting profession to greater connectivity and transparency and as a result, greater stakeholder confidence (Sidhu 2019).

89.5.2 Financial Misstatements/Misreporting/Frauds

Predicting financial misstatements or frauds is another direction for the research of machine learning's applications in auditing. One of the early research is conducted by Cecchini et al. (2010b). Instead of using financial ratios, they analyze basic financial data to detect management frauds with support vector machine. A finance-specific kernel that constructs features from existing literature is developed to increase the power of the learning. The developed model outperforms the traditional fraud prediction models and is able to successfully identify 80% of the frauds and 90.6% of the nonfraudulent observations.

With a dataset consisting of 51 fraud observations and 15,934 nonfraud cases between 1998 and 2005, Perols (2011) compares the performance of six machine learning algorithms (i.e., artificial neural networks, support vector machine, decision tree, logistic regression, bagging, and Stacking) in detecting financial statement fraud. The results show that logistic regression and support vector machines outperform other algorithms. The research finds that six predictors are consistently important for all algorithms: auditor turnover, total discretionary accruals, Big 4 auditor, accounts receivable, meeting or beating analyst forecasts, and unexpected employee productivity. Based on the Perols (2011) dataset, Perols et al. (2017) increase the number of explanatory variables to 109. The variables are primarily derived from prior literature (i.e., Cecchini et al. 2010b; Dechow et al. 2011; Perols 2011). More than 10,000 prediction models are developed with three data analytics preprocessing methods. Their results indicate that two of these three methods improve fraud prediction accuracy by approximately 10%.

Recently, Bao et al. (2020), following Cecchini et al. (2010), use raw financial data items from financial statements of US public companies between 1991 and 2008. They apply ensemble learning to predict frauds and find that an ensemble learning model produces more accurate predictions relative to the two benchmark models (i.e., Dechow et al. 2011 and Cecchini et al. 2010b). Bertomeu et al. (2020) use the Audit Analytics Advanced Non-Reliance Restatement database to collect samples of material misstatements from 2001 to 2014. In total, the sample in this literature consists 3,599 misstatement observations and 54,354 Compustat observations, with financial ratios as predictors as well as a wide range of other variables from governance and auditing datasets, etc. The main machine learning algorithm is gradient boosted regression tree (GBRT), which is a regression method that explains the probabilities of misstatements. It finds that, as opposed to traditional methods, machine learning approaches improve the prediction accuracy of out-of-sample test. The literature also employs a number of alternative methods and documents that the classification algorithms generally perform well at predicting misstatements especially when the sample size is big enough.

On the same topic, Brown et al. (2020) focus on the thematic content of financial reports. They utilize Bayesian topic modeling algorithm to identify and quantify the topic content of 10-K filings spanning from 1994 to 2012 and document that this information is useful for predicting misstatements. This literature extends the

research stream by exploring the semantic meaning of management disclosure and how disclosure themes point to misreporting. Similar work includes Cecchini et al. (2010a), Goel et al. (2010), and Purda and Skillicorn (2015). They all use support vector machine to classify misstatements. Thanks to this technique, there is no need to require predefined word dictionaries and the results show that their models provide better prediction performance than prediction models and those based on quantitative financial measures.

89.5.3 Audit Quality

Audit quality is a complex problem and there are a variety of measures used by existing literature which generate mix results. Hu et al. (2020) propose a machine learning approach to measure audit quality by constructing “surprise” scores based on traditional proxies of audit quality: audit adjustments, nonclean audit opinions, and restatements. They assert that the proposed measure of audit quality is capable of capturing unobservable factors that drive audit quality because it is the difference between the predicted probability of an event (e.g., the issuance of a modified audit opinion) and the actual value. It is the first archival research applying multiple types of audit adjustments and utilizing a comprehensive characteristics of audit firms and individual audit partners in addition to financial variables for audit quality measuring. They demonstrate the application of the proposed audit quality measures for auditing research by examining the relationship between the constructed measure and audit failure.

Yang et al. (2020) implement the double machine learning (DML) approach that relies on machine learning methods to examine the Big N audit quality effect. They demonstrate that the gradient boosting method outperforms other machine learning algorithms (e.g., decision tree, random forest, and support vector machine) for DML. They apply the DML approach with gradient boosting to estimate the Big N effect and evidence that Big N auditors have a positive effect on audit quality.

Different from prior studies, Hunt et al. (2021) employ machine learning techniques to predict the likelihood of auditor switches using data samples from 2002 through 2016. They compare the accuracy of predictions generated by multiple machine learning models including gradient boosting, random forest, artificial neural networks, support vector machines, and logistic regression. The results show that gradient boosting consistently outperforms the other algorithms. They use the probability of switching auditors estimated by the gradient boosting model and investigate whether increased probability of auditor switches affect audit quality as proxied by misstatements and abnormal accruals. This chapter finds that lower audit quality among companies that are more likely to switch auditors but remain with their incumbent auditor.

89.6 Machine Learning Tools

There is a series of publicly available tools that effortlessly implement machine learning algorithms.

89.6.1 Scikit-Learn Module in Python

Aiming to bring machine learning to nonspecialists in programming, Scikit-learn is a python module that incorporates various machine learning techniques ranging from data preprocessing to supervised and unsupervised learning (Pedregosa et al. 2011). Examples, source codes, and documentation of specific learning algorithms can be found on the official website <https://scikit-learn.org/stable/>. When employing Scikit-learn to perform supervised learning tasks, the core method is fit, referring to training the machine learning models (e.g., decision tree classifier) based on data samples from the train set. After training, the predict method can infer the target labels or numerical outcomes of previously unseen data points. Concerning the unsupervised learning algorithms, such as K-means clustering, the transform method can be exploited to acquire the modified versions of data sets. Furthermore, Scikit-learn offers the cross-validation method⁹ to help users tune the parameters of machine learning algorithms and evaluate models' performance by using task-relevant metrics.

89.6.2 PyOD Module in Python

Recently, some outlier detection methods, such as isolation forest, local outlier factor, and one-class SVM, are incorporated into the audit process. These techniques empower auditors to analyze the entire ledger/transaction sets and find high-risk areas that deserve further investigation (Analyticsai 2020). PyOD is an open-source Python toolkit for identifying anomalies existing in multivariate data (Zhao et al. 2019). It embraces a wide range of outlier detection algorithms, from traditional K-nearest neighbors to the newly invented AutoEncoder. Researchers may use this toolbox to explore which set of outlier detection methods fits in the audit context.

89.6.3 TensorFlow Platform

Invented by the Google Brain team, TensorFlow is an open-source library that supports fast numerical computations. This toolkit can be applied in the construction of traditional machine learning and deep learning models.¹⁰

⁹For details of the cross-validation method, check the website: https://scikit-learn.org/stable/modules/cross_validation.html

¹⁰More details are presented on the website: <https://www.tensorflow.org/>

89.7 Conclusions: Perspectives for Future Research

The research on the applications of machine learning in accounting and auditing is still evolving. There are numerous promising areas that have not been explored. One of these areas is using machine learning to estimate uncollectable bad debts. The current method to determine the allowance for doubtful accounts is to estimate the percentage of receivable based on the age of outstanding accounts receivables or to use a percentage based on prior years' actual uncollectible accounts to prior years' credit sales percentage of sales. We are expecting a machine learning model trained with history data that describing the business, the auditor, the clients, as well as the economic and social environment to make better predictions.

While the vast majority of existing research is centered on the problem of financial misstatements or frauds, other topics are of great potential for the implementation of machine learning to improve the quality of financial information. Examples of those topics include material internal control weaknesses, real earnings management (i.e., management operational activities to change reported earnings in a particular direction by overproducing inventory to lower the cost of goods sold or cutting discretionary expenses to improve reported margins (Huang and Sun 2017)), and accrual-based earnings management deserve investigation (i.e., management actions of altering the accounting methods, such as the depreciation method for fixed assets, or estimates, such as the estimate for provision for doubtful accounts, used when presenting a given transaction in the financial statements (Zang 2012)).

We are now in the era of big data. Audit methodologies and practices need to be aligned with the development of cutting-edge technologies and make sufficient use of the exogenous data from a variety of sources to improve the effectiveness and efficiency of accounting work. Exogenous data refers to "data from third-party sources rather than internal accounting records of corporate entities (Brown-Liburd et al. 2019)." The exogenous data, such as posts on social media, location data collected from Internet of Things, weather data, and google search data, can be analyzed with machine learning techniques to extract informative data like the sentiment, topic, entity, and keywords. Future research can utilize those exogenous data as a supplement to endogenous financial data to develop prediction model to support audit decision-making.

To make the results understandable and transparent to humans, future research needs to take effort to explore the interpretability of "black box" models of machine learning. While feature importance (i.e., a technique for determining which variable(s) contribute(s) most strongly to a model's output) has been used frequently in prior studies, other techniques are rarely used in accounting and auditing literature to interpret the results. For instance, accumulated local effects (ALE) (Apley and Zhu 2020) provides global explanations for machine learning models by using an ALE plot to describe how features influence the prediction of the model on average. The Shapley value¹¹ is a concept in cooperative game

¹¹Shapley, Lloyd S. (August 21, 1951). "Notes on the n-Person Game -- II: The Value of an n-Person Game" (PDF). Santa Monica, Calif.: RAND Corporation.

theory,¹² which assigns a unique distribution (among the players) of a total surplus generated by the coalition of all players. When the Shapley value is used in machine learning, we assume that each value of the input variable for a given sample is a player in a cooperative game where the prediction is the payout. So Shapley value helps us explain how much each value of the input variable has contributed to the prediction compared to the average prediction (Roth 1988). Local interpretable model-agnostic explanations (LIME) can be applied to any machine learning model to help us understand the model by training local surrogate models¹³ to explain predictions for each single data sample. Specifically, it modifies a single data sample by adjusting the feature values and observes the resulting impact on the prediction.

Besides archival research, behavior studies can be conducted to explore questions like how the decision-making of humans affected by the implementation of machine learning technologies, to what extent human auditors rely on the results generated by machine learning models, what the performance of machine learning techniques is when they are adopted to conduct structured and repetitive tasks as opposed to unstructured tasks that need higher level of professional judgment, and how investors or other stakeholders perceive and evaluate the quality of financial reports if they know the reports are prepared/audited with the help of machine learning techniques.

Future research could discuss that the increased use of machine learning and other AI techniques will require human accountants to identify and manage risks with this technology through systematic governance and internal controls. It is also very important that future research examines the inherent bias in machine learning and how it can be addressed in practice. The bias comes from multiple sources, such as the data-driven bias due to the flaws or skewness in the data we collected for model training purpose and the interaction bias between human and machine as machines learn the biases of the people who train them (Hammond 2016). Furthermore, providing solutions to data privacy and security issues is another direction for future research. There is still a long way to go for machine learning applications in the context of accounting auditing. Scholars, practitioners, and policy-makers need to cooperate more closely to foster the development of the accounting profession with the help of innovative technology and provide guidance for the best use of machine learning to make accounting and auditing more effective and efficient.

¹²Read “Cooperative game theory assumes that groups of players, called coalitions, are the primary units of decision-making, and may enforce cooperative behavior.” (Choudhary 2019). <https://www.analyticsvidhya.com/blog/2019/11/shapley-value-machine-learning-interpretability-game-theory/>

¹³Local surrogate models are interpretable models that are used to explain individual predictions of black box machine learning models.

References

- Adadi, A., and M. Berrada. 2018. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* 6: 52138–52160.
- Agrawal, R., T. Imieliński, and A. Swami. 1993. Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD international conference on Management of data, pp 207–216.
- Alpaydin, E. 2020. *Introduction to machine learning*. Cham: MIT Press.
- Altman, E.I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23: 589–609.
- Analyticsai, C. 2020. *AnalyticsAI for every engagement* [Online]. Available: <https://www.caseware.com/us/analyticsai>. Accessed.
- Anand, V., R. Brunner, K. Ikegwu, and T. Sougiannis. 2019. Predicting profitability using machine learning. Available at SSRN 3466478.
- Anthony, M., and P.L. Bartlett. 2009. *Neural network learning: Theoretical foundations*. Cambridge: Cambridge University Press.
- Apley, D.W. 2016. Visualizing the effects of predictor variables in black box supervised learning models. arXiv preprint arXiv:1612.08468.
- Apley, D.W., and J. Zhu. 2020. Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 82 (4): 1059–1086.
- Bao, Y., and A. Datta. 2014. Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science* 60: 1371–1391.
- Bao, Y., B. Ke, B. Li, Y.J. Yu, and J. Zhang. 2020. Detecting accounting fraud in publicly traded US firms using a machine learning approach. *Journal of Accounting Research* 58 (1): 199–235.
- Barboza, F., H. Kimura, and E. Altman. 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications* 83: 405–417.
- Barth, Mary E. and Li, Ken and McClure, Charles. 2021. Evolution in Value Relevance of Accounting Information. Stanford University Graduate School of Business Research Paper No. 17-24, Available at SSRN: <https://ssrn.com/abstract=2933197> or <https://doi.org/10.2139/ssrn.2933197>.
- Beneish, M.D. 1999. The detection of earnings manipulation. *Financial Analysts Journal* 55: 24–36.
- Bertomeu, J. 2020. Machine learning improves accounting: Discussion, implementation and research opportunities. *Review of Accounting Studies* 25: 1135–1155.
- Bertomeu, J., E. Cheynel, E. Floyd, and W. Pan. 2020. Using machine learning to detect misstatements. *Review of Accounting Studies* 26: 1–52.
- Bishop, C.M. 2006. *Pattern recognition and machine learning*. Springer.
- Breiman, L. 1996. Bagging predictors. *Machine Learning* 24: 123–140.
- Brown, N.C., R.M. Crowley, and W.B. Elliott. 2020. What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research* 58 (1): 237–291.
- Brown-Liburd, H., A. Cheong, M.A. Vasarhelyi, and X. Wang. 2019. Measuring with exogenous data (MED), and government economic monitoring (GEM). *Journal of Emerging Technologies in Accounting*. 16 (1): 1–19.
- Bzdok, D., N. Altman, and M. Krzywinski. 2018. Points of significance: Statistics versus machine learning. *Nature Methods* 15 (4): 233–234. <https://www.nature.com/articles/nmeth.4642.pdf?origin=ppub>.
- Carton, R.B., and C.W. Hofer. 2006. *Measuring organizational performance: Metrics for entrepreneurship and strategic management research*. Edward Elgar Publishing.
- Cecchini, M., H. Aytug, G.J. Koehler, and P. Pathak. 2010a. Making words work: Using financial text as a predictor of financial events. *Decision Support Systems* 50: 164–175.
- Cecchini, A., et al. 2010b. Detecting management fraud in public companies. *Management Science* 56 (7): 1146–1160. <https://doi.org/10.1287/mnsc.1100.1174>.

- Chen, M.-S., J. Han, and P.S. Yu. 1996. Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering* 8: 866–883.
- Cho, S., M.A. Vasarhelyi, T. Sun, and C. Zhang. 2020. *Learning from machine learning in accounting and assurance*. Journal of Emerging Technologies in Accounting.
- Chollet, F. 2017. *Deep learning with python*. Shelter Island: Manning Publications Company.
- Choudhary, A. 2019. Analytics vidhya. A unique method for machine learning interpretability: Game theory & shapley values. <https://www.analyticsvidhya.com/blog/2019/11/shapley-value-machine-learning-interpretability-game-theory/>.
- Dechow, P.M., and I.D. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* 77: 35–59.
- Dechow, P.M., W. Ge, C.R. Larson, and R.G. Sloan. 2011. Predicting material accounting misstatements. *Contemporary Accounting Research* 28 (1): 17–82.
- Dietterich, T.G. 2002. Ensemble learning. In *The handbook of brain theory and neural networks*, vol. 2, 110–125. Cambridge, MA: MIT Press.
- Ding, K., B. Lev, X. Peng, T. Sun, and M.A. Vasarhelyi. 2020. Machine learning improves accounting estimates: Evidence from insurance payments. Available at SSRN 3253220.
- Expert.ai. 2020. What is machine learning? A definition. <https://www.expert.ai/blog/machine-learning-definition/>
- Foote, K.D. 2019. A brief history of machine learning. Data Topics. Dataversity. <https://www.dataversity.net/a-brief-history-of-machine-learning/>
- Frankel, R., J. Jennings, and J. Lee. 2016. Using unstructured and qualitative disclosures to explain accruals. *Journal of Accounting and Economics* 62: 209–227.
- Ghahramani, Z. 2015. Probabilistic machine learning and artificial intelligence. *Nature* 521 (7553): 452–459. <https://www.repository.cam.ac.uk/bitstream/handle/1810/248538/Ghahramani%202015%20Nature.pdf;jsessionid=3DB2D31FFA80196A97AEEBECB06FEF42?sequence=1>.
- Goel, S., J. Gangolly, S.R. Faerman, and O. Uzuner. 2010. Can linguistic predictors detect fraudulent financial filings. *Journal of Emerging Technologies in Accounting*. 7: 25–46.
- Hammond, K. 2016. 5 unexpected sources of bias in artificial intelligence. Available at: <https://techcrunch.com/2016/12/10/5-unexpected-sources-of-bias-in-artificial-intelligence/>
- Healthcare.ai. 2020. Machine learning versus statistics: When to use each. Data Science Blog. <https://healthcare.ai/machine-learning-versus-statistics-use/>
- Hebb, D.O. 1949. *The organization of behavior: A neuropsychological theory*. New York, London: J. Wiley, Chapman & Hall. http://s-f-walker.org.uk/pubsebooks/pdfs/The_Organization_of_Behavior-Donald_O._Hebb.pdf.
- Heller, M. 2019. *Machine learning algorithms explained* [Online]. Available: <https://www.infoworld.com/article/3394399/machine-learning-algorithms-explained.html>. Accessed.
- Huang, X.S., and L. Sun. 2017. Managerial ability and real earnings management. *Advances in Accounting* 39: 91–104.
- Huang, A.H., A.Y. Zang, and R. Zheng. 2014. Evidence on the information content of text in analyst reports. *The Accounting Review* 89: 2151–2180.
- Hu, H., T. Sun, M.A. Vasarhelyi, and M. Zhang. 2020. A Machine Learning Approach of Measuring Audit Quality: Evidence From China. Available at SSRN 3732563.
- Huang, A.H., R. Leheavy, A.Y. Zang, and R. Zheng. 2018. Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science* 64: 2833–2855.
- Hunt, J.O., D.M. Rosser, and S.P. Rowe. 2021. Using machine learning to predict auditor switches: How the likelihood of switching affects audit quality among non-switching clients. *Journal of Accounting and Public Policy* 40(5): p.106785.
- Khalid, S., T. Khalil, and S. Nasreen. 2014. A survey of feature selection and feature extraction techniques in machine learning. 2014 Science and information conference. IEEE, 372–378.
- Kim, H.S., and S.Y. Sohn. 2010. Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research* 201: 838–846.
- Kober, J., J.A. Bagnell, and J. Peters. 2013. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research* 32: 1238–1274.

- Lecun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. *Nature* 521: 436–444.
- Lefkowitz, M. 2019. Professor's perceptron paved the way for AI: 60 years too soon. Cornell Chronicle. <https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon>
- Li, F. 2010. The information content of forward-looking statements in corporate filings – A naïve Bayesian machine learning approach. *Journal of Accounting Research* 48: 1049–1102.
- Odom, M.D., and R. Sharda. 1990. A neural network model for bankruptcy prediction. 1990 IJCNN International Joint Conference on neural networks. IEEE, 163–168.
- Ohlson, J.A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18: 109–131.
- Olson, D.L., D. Delen, and Y. Meng. 2012. Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems* 52: 464–473.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg. 2011. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research* 12: 2825–2830.
- Perols, J. 2011. Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing* 30 (2): 19–50.
- Perols, J.L., R.M. Bowen, C. Zimmermann, and B. Samba. 2017. Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review* 92 (2): 221–245.
- Platt, H.D., M.B. Platt, and J.G. Pedersen. 1994. Bankruptcy discrimination with real variables. *Journal of Business Finance & Accounting* 21: 491–510.
- Provalis Research. 2017. Blogs on Text Analytics: A Brief History of Machine Learning. <https://provalisresearch.com/blog/brief-history-machine-learning/>.
- Purda, L., and D. Skillicorn. 2015. Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research* 32: 1193–1223.
- Rosenblatt, F. 1957. The perceptron: A perceiving and recognizing automation (Project Para). <https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf>
- Roth, Alvin E., ed. 1988. *The Shapley value: Essays in honor of Lloyd S. Shapley*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511528446>. ISBN 0-521-36177-X.
- Sallab, A.E., M. Abdou, E. Perot, and S. Yogamani. 2017. Deep reinforcement learning framework for autonomous driving. *Electronic Imaging* 2017: 70–76.
- Shalev-Shwartz, S., and S. Ben-David. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge: Cambridge University Press.
- Shaw, R. 2017. *Top 10 machine learning algorithms for beginners* [Online]. KDnuggets. Available: <https://www.kdnuggets.com/2017/10/top-10-machine-learning-algorithms-beginners.html>. Accessed.
- Shin, K.-S., T.S. Lee, and H.-J. Kim. 2005. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications* 28: 127–135.
- Sidhu, H. 2019. How audit digitization reflects a transformative age. Available at: https://www.ey.com/en_gl/digital-audit/auditdigitization-transformative-age
- Sun, T. 2019. Applying deep learning to audit procedures: An illustrative framework. *Accounting Horizons* 33 (3): 89–109.
- Sutton, R.S., and A.G. Barto. 2018. *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
- Tsai, C.-F., Y.-F. Hsu, and D.C. Yen. 2014. A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing* 24: 977–984.
- Van Den Bogaerd, M., and W. Aerts. 2011. Applying machine learning in accounting research. *Expert Systems with Applications* 38: 13414–13424.
- Van Der Maaten, L., E. Postma, and J. Van Den Herik. 2009. Dimensionality reduction: A comparative. *Journal of Machine Learning Research* 10: 13.
- Wiederhold Gio, John McCarthy, and Ed Feigenbaum. 1990. “Memorial resolution: Arthur L. Samuel” (PDF). Stanford University Historical Society. Archived from the original (PDF)

- on 26 May 2011. Retrieved April 29, 2011. <https://web.archive.org/web/20110526195107/http://histsoc.stanford.edu/pdfmem/SamuelA.pdf>
- Yang, Z., M.B. Platt, and H.D. Platt. 1999. Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research* 44: 67–74.
- Yang, J.C., H.C. Chuang, and C.M. Kuan. 2020. Double machine learning with gradient boosting and its application to the Big N audit quality effect. *Journal of Econometrics* 216: 268–283.
- Zang, A.Y. 2012. Evidence on the trade-off between real activities manipulation and accrual-based earnings management. *The Accounting Review* 87 (2): 675–703.
- Zhao, Q., and S.S. Bhowmick. 2003. *Association rule mining: A survey*. Vol. 135. Singapore: Nanyang Technological University.
- Zhao, Y., Z. Nasrullah, and Z. Li. 2019. Pyod: A python toolbox for scalable outlier detection. *arXiv preprint arXiv:1901.01588*.
- Zhou, Z.-H. 2009. Ensemble learning. In *Encyclopedia of biometrics*, vol. 1, 270–273. New York: Springer.

Internal Capital Budgeting and Allocation in Financial Firms

90

Woo-Young Kang

Abstract

Internal capital budgeting for financial firms differs from that for nonfinancial firms. Financial firms are highly leveraged, risk-sensitive, and possess limited equity capital to cover their illiquid risks compared to nonfinancial firms. Thus, some issues regarding internal capital budgeting – such as type of firm, financial shock, and capital budgeting methodologies – are somewhat different in case of financial and nonfinancial firms, while almost all their other characteristics including agency problem, optimal timing for capital budgeting, relative dependence on the external capital market, firm boundary, and market type remain near identical. Internal capital budgeting for financial firms aims to maximize their risk-adjusted performance measured by risk-adjusted return on capital (RAROC), return on risk-adjusted capital (RORAC), risk-adjusted return on risk-adjusted capital (RARORAC) and so on, leading to the maximization of shareholder value. On the other hand, internal capital budgeting for nonfinancial firms is largely affected by the traditional standard discounted cash flow (DCF) method. Financial firms are required to reflect regulatory capital requirements in their internal capital budgeting process, and using RORAC or RARORAC performance measures, which emphasize the risk component of capital, can be quite helpful for this.

Keywords

Internal Capital Market · External Capital Market · Financial Firms · Nonfinancial Firms · Capital Budgeting · Economic Capital · Financial Regulation

W.-Y. Kang (✉)

Department of Economics and Finance, Brunel University London, Uxbridge, UK

e-mail: woo-young.kang@brunel.ac.uk

© Crown 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_92

2117

JEL ClassificationD82 · E22 · G20 · G21 · G23 · G28 · G31

90.1 Backgrounds of Internal Capital Budgeting

In the economy, a firm has access to two types of markets as its source of financing – external and internal. When a firm engages with the external capital market, it interacts with external capital suppliers such as the stock market, the bond market, financial intermediaries, etc. On the other hand, a firm interacting with the internal capital market tries to conduct its business with the headquarter, which allocates its limited capital to the business units it owns. Therefore, both the external and the internal capital market can be a source of capital for a firm and its business units (Gertner et al. 1994). However, there are certain problems that firms need to solve in order to efficiently utilize the internal capital market for capital budgeting, and a financial firm's approach to do so is very different from and more specific than that of a nonfinancial firm. This chapter aims to analyze the issues and approaches regarding internal capital budgeting for financial firms in comparison with those for non-financial or general firms.

Firms can create value through capital budgeting in the internal capital market. A firm's headquarters selects optimal investment projects from the various divisions and reallocates capital by moving it from relatively low-return projects to high-return projects or by using one division's assets as collateral to fund another division based on various criteria. These criteria include industry or business unit characteristics (e.g., Bardolet et al. 2011), desires for growth (e.g., Arrfelt et al. 2015), opportunities for creating synergies (e.g., Bower and Gilbert 2005), and need to reduce corporate risk to increase returns (e.g., Liebeskind 2000). There are three major motives behind the internal capital allocation according to Busenbark et al. (2017). First is the winner picking motive (e.g., Stein 1997), which selects and allocates capital to business units with the best performance prospects and maximize the overall firm growth (e.g., Arrfelt et al. 2015; Duchin and Sosyura 2013; Wulf 2009). Risk mitigation (e.g., Matvos and Seru 2014) is the second motive, which involves diversification. Corporate managers can allocate capital to their related or unrelated business units such that their firm can limit exposure to exogenous shocks to its business unit and/or maximize the variety of its product offerings (e.g., Lang and Stulz 1994; Liebeskind 2000; Shin and Stulz 1998). The third motive is the exploitation of synergies through leveraging and sharing of capabilities and assets across business units (e.g., Bower and Gilbert 2005). In other words, internal capital allocation may share, lend, and leverage resources, capabilities, and expertise among business units with disparate characteristics (e.g., Cremers et al. 2011; Williamson 1991).

Capital budgeting in the internal capital market depends on both the absolute and relative merits of the investment projects. For instance, if a company has two unrelated projects and the relative merit of investing in one of them is higher, this

implies that the investment in the other project, despite the fact that its net present value (NPV) may be positive, should decrease. Although important, the NPV is not the only relevant benchmark for internal capital budgeting decisions. There are also other important criteria that determine the budgeting decision, including the project's strategic fit, its cash flow structure, market share, past performance, and the business unit CEOs reputation, confidence, and gut feel behind the project (Graham et al. 2015; Glaser et al. 2013). The size and scope of the internal capital market also affects its value creation when there is a trade-off. If the size and scope of the internal capital market increases, the absolute amount of investment will increase as well. Consequently, financial constraints will loosen and a firm can raise more capital. However, a side effect of this is that the monitoring costs to the headquarters also increase and it becomes more difficult to select an optimal project (Stein 1997).

The capital budgeting of the internal capital markets has both bright and dark sides. The bright side views the internal capital allocation as a result of pooling internally generated cash flows optimally distributed to business units as funds determined by their investment prospects (Weston 1970; Williamson 1975; Matsusaka and Nanda 2002; Maksimovic and Phillips 2002), a necessary condition to make the right investments over time (Lovallo et al. 2020). This winner-picking process creates value by making more allocations to units with better investment opportunities (Stein 1997). On the other hand, the dark side views the hierarchy within a firm as disproportionately impacting internal policies and leading to inefficient capital budgeting decisions (Coase 1937). The business unit or divisions run by more powerful or better connected managers may secure more internal funding than their justified amount based on their investment opportunities and outlooks (Meyer et al. 1992; Scharfstein and Stein 2000; Rajan et al. 2000; Wulf 2009). These managers may prefer larger capital allocations for rent-seeking or empire-building purposes and exert more influence to secure more funds, which are costly and induce inefficient investment decisions for the firm (Glaser et al. 2013).

90.2 Issues Related to Internal Capital Budgeting

I start by discussing issues relevant to capital budgeting in the internal capital market derived from the extant literature in the context of general firms. These issues include the agency problem, optimal timing for capital budgeting, relative dependence on the external capital market, firm type, firm boundary, market type, financial shock, capital budgeting methodologies, etc. Many of these issues also relate to financial firms' internal capital budgeting, although they may still need some adjustments on account of their highly leveraged and risk-sensitive characteristics.

90.2.1 The Agency Problem

The agency problem is an internal capital budgeting issue that mainly comprises managerial issues and information asymmetry.

90.2.1.1 Managerial Issues

Managerial issues are a major part of the agency problem. In the internal capital market, the CEO's opinion of the divisional manager is the second most influential factor after the NPV rule (Graham et al. 2015). In other words, the power balance and connection between the CEO and divisional managers can distort the internal capital budgeting. Specifically, managers who have good relationships with their bosses receive more capital through the "appointment channel" or the "capital allocation channel." The appointment channel designates connected managers to divisions with more capital and, following this, connected managers receive more capital than other nonconnected ones through the capital allocation channel. The effect of the capital allocation channel is approximately two times greater than that of the appointment channel. However, these managers' connections with the CFOs or board members have a weaker effect on internal capital budgeting compared to those who have connections with CEOs (Duchin and Sosyura 2013). The misalignment of investment incentives between divisional managers and headquarters may result in inefficient investments, which may be either over or under investments (Lamont 1997; Shin and Stulz 1998; Rajan et al. 2000; Hund et al. 2010; Ozbas and Scharfstein 2010; Roper and Ruckes 2012). CEO duality can also be a source of incentive misalignment between the role as a CEO and a chair of the board. This may result in inefficient investments by allocating more capital to business units with low growth prospects over units with high growth prospects (Aktas et al. 2019), and CEO's equity-based compensation can ameliorate this negative impact (Datta et al. 2009; Aktas et al. 2019). The investment inefficiencies resulting from managerial power and connections occur especially in case of a firm-wide financial slack (Glaser et al. 2013).

There are also other agency problems related to managerial issues. These include insiders' voting rights exceeding their cash flow rights (Hoechle et al. 2012), managers' rent-seeking activities (Duchin and Sosyura 2013), empire building tendencies (Xuan 2009; Hoechle et al. 2012), over-confidence (Hoechle et al. 2012) or voluntary risk-adversity (Stoughton and Zechner 2007) from different risk appetites, the relationship between managers' incentives and compensation (Huang et al. 2012), managers' behavioral biases or cognitive limitations (Bardolet et al. 2011; Arrfelt et al. 2013), sociopolitical forces on managers within the organization (Bernardo et al. 2006; Xuan 2009; Glaser et al. 2013; Hoang and Ruckes 2015), etc.

90.2.1.2 Information Asymmetry

Information asymmetry is another major part of the agency problem. It is mostly caused by divisional managers' monopoly over business information such as changes in customer demands, the likelihood of market entry or exit, quality or pricing adjustments of competitors, and new technologies for product development. Information asymmetry among divisions can generate a coordination problem since all the divisions compete to receive capital budgets from headquarters as early as possible, simultaneously. It also affects the strategic behavior of headquarters when moving resources from divisions with a high degree of information asymmetry to divisions with more information transparency (Roper and Ruckes 2012). When external financial shock negatively causes some high-ability agents (or divisional managers) to

abandon their investment projects, other low-ability agents use this opportunity as shelter to hide their private information by delaying abandoning their poor investment projects until the high-ability agents abandon theirs (Grenadier et al. 2014).

90.2.2 Optimal Timing for Capital Budgeting

The optimal timing for capital budgeting has been discussed mostly in the context of real options. This includes competition among divisional managers for securing capital allocation, past financial shocks, investment-signaling incentives, and others. Competition among divisional managers affects the optimal time for capital budgeting. Divisional managers compete with each other to secure capital allocation as soon as possible, since headquarters would have a limited budget that all the managers would strive to secure as soon as possible, simultaneously. This results in a coordination problem resulting from uncertainty in investment opportunities and the actions of other divisions. Under conditions of such severe financial constraints, it is optimal to reduce the probability of significant capital allocation in the second period if a significant capital allocation takes place in the first period – this is known as the dynamic rationing policy (Roper and Ruckes 2012).

The traditional approach to achieve optimal timing in capital allocation for investment demonstrates the importance of “the option to wait” in the real-option context due to uncertainty regarding future shocks to cash flows. However, lessons from past financial shocks, called “option to learn,” is another way to determine the timing of investment. When a firm faces a financial shock and it is not yet clear whether it is temporary or permanent, it will want to wait to learn more about the nature of the shock. If the firm continues to see high cash flow in the next period, the past shock is likely to reflect a positive fundamental change. On the other hand, if it witnesses a decrease in cash flow, this past shock is more likely to be the result of temporary non-fundamental fluctuations (Grenadier and Malenko 2010).

The investment signaling incentives of internal managers also affect investment timing. It is to fool the outside investors and take advantage of the actual investment timing to maximize the inside managers’ profits. Internal managers display four different kinds of behavior depending on the scenario. First, when inside managers care about both the long-term performance of the company as well as its short-term stock price, they tend to invest early in order to fool the market into overestimating their investment project’s NPV and raising their stock price. Second, when a manager is a venture capitalist with a short track record, he or she will struggle to take the company public early, so as to establish a reputation and raise more money to fund the business. Third, if a manager attempts to take some of the cash flow from investment as his or her private benefit, he or she will delay the investment to signal that the investment project’s NPV is lower than it really is. In this manner, a manager can divert some allocated investments from headquarters for his or her own personal consumption. Fourth, inside managers have all the information on their investment projects, so they will delay the decision to invest to fool their rival firms into underestimating their investment projects (Grenadier and Malenko 2011).

90.2.3 Relative Dependence on the External Capital Market

The relative dependence on the external market affects the capital budgeting strategy of a company's headquarters as well as its various divisions. When the external capital market has more abundant cash flow than the internal capital market, a company's divisions seek more financing from the external capital market. As a result, the company's headquarters will reduce its capital charge. The divisional managers will try to take advantage of this reduced capital charge by receiving capital allocation from headquarters. Accordingly, these managers will exaggerate the returns of their investment opportunities to headquarters to receive more capital. This creates distortions in investment information, leading to inefficient capital allocation.

In contrast, when divisional managers seek more financing from the internal capital market than from the external one, headquarters will raise the capital charge to extract surplus from the divisional managers. In response, these managers will under-report their investment returns to avoid high capital charges, which perverts the investment information, leading to inefficient capital allocation (Stoughton and Zechner 2007).

Specifically, for financial firms, a contingent convertible (CoCo) bond is a very recent example of an external capital market source for financing, especially for financial firms. A CoCo bond is a fixed-income security that converts into equity or writes-down its principal when the pre-determined lower capital ratio threshold is breached (De Spiegeleer et al. 2014). Therefore, it is a source of equity for financial firms to protect themselves from bankruptcy and satisfy the financial regulation requirements at the cost of their relatively high coupon payments to their investors.

90.2.4 Type of Firm

A firm's capital budgeting process depends on what type of firm it is, e.g., public/private, listed/unlisted, levered/unlevered. In countries with well-developed stock markets, the capital budgeting of public listed firms is more effective, since their investments are more sensitive to growth opportunities compared to the investments of public unlisted and private firms. In these markets, public listed firms can benefit from investments, which may not be possible to undertake if they were public unlisted or private firms (Mortal and Reisel 2013). The amount of a firm's risky debt also affects its internal capital budgeting. A levered firm invests earlier than an unlevered firm, because the option to wait becomes less valuable as a higher debt level leads to a higher default probability in the absence of wealth transfers among the claim holders (Lyandres and Zhdanov 2010).

90.2.5 Firm Boundary

A firm's internal capital budgeting can be affected by its internal and external boundaries. A firm's internal research and development (R&D) activities in particular face significant information asymmetry between its headquarters and its various

divisions. The managers involved in the R&D process have significant soft information, which can be difficult to communicate. Besides, they may conceal unfavorable information that could possibly emerge during the research stage, resulting in inefficient internal capital allocation (Scherer 1999; Seru 2014). This can happen if the costs of providing incentives inside the firm are greater than the benefits of innovative research. However, the problem can be alleviated if the firm shifts its internal R&D activity to the external boundaries of the firm (Seru 2014).

90.2.6 Type of Market

A firm's capital budgeting can be affected by its market type – competitive or concentrated. In a competitive market, a firm must follow the critical rule of survival, which is to invest as early as possible to avoid losing the investment opportunity to its competitors. It is optimal to invest when the value of the additional capacity is larger than the investment cost by an amount that corresponds to the value of the option to wait. However, in a competitive market, the value of the option to wait diminishes more than the difference between the value of the additional capacity and the investment cost, which discourages waiting for investments. In a concentrated market, on the other hand, a firm will have higher values both in case of the option to wait for investment and investment in additional capacity compared to a firm in a competitive market. This relatively encourages waiting for investment than in a competitive market. A firm in a concentrated market gains a higher value from the investment of additional capacity compared to the existing value of future cash flows from its installed capacity (Aguerrevere 2009).

90.2.7 Financial Shock

Internal capital budgeting interacts with financial shock. For instance, changes in stock price that reflect financial shock can signal changes in investment opportunities in a firm, which in turn affects its capital budgeting across its investment projects. These investments become more sensitive to financial shock when the firms are more levered, have less cash and cash flow and have lower dividends (Ovtchinnikov and McConnell 2009). Financial firms facing a financial shock or a banking crisis can see a negative impact on their internal capital budgeting, which is worse in developed countries. This is because financial firms in developed countries will move their intangible assets to firms in less developed countries with globalization activities and cross-border characteristics (Fernández et al. 2013). Conversely, internal capital budgeting can also help mitigate the negative effect of a financial shock. However, this process involves a significant amount of corporate socialism. Too little capital is allocated to stronger divisions and too much to weak divisions, in relation to their actual productivities (Matvos and Seru 2014).

90.2.8 Capital Budgeting Methodologies

Most companies rely on traditional standard discounted cash flow (DCF) calculations to make investment decisions. DCF is also a necessary input for real option valuation (McDonald 2006). The real options approach to capital budgeting involves actively managing projects to enhance the value of capital investments (Mittendorf 2004) and greatly simplifies difficult valuation problems (McDonald 2006). There have also been other capital budgeting methods reflecting agency problems (e.g., as discussed in Sect. 90.2.1), external financing (e.g., Almeida and Wolfenzon 2005), etc.

However, financial firms are highly leveraged since they issue equity infrequently but have continuous access to debt capital (Stoughton and Zechner 2007) compared to nonfinancial firms. Thus, financial firms require capital budgeting practices to be linked with risk management functions (Froot and Stein 1998). In addition, as their businesses are credit sensitive and largely depend on default risks, risk capital (i.e., equity capital) is crucial for them (Perold 2005) and should be allocated (Erel et al. 2015). There are several major risk capital allocation methods, such as the Euler allocation principle, the Lagrange method, capital-loss distance minimization, etc., all of which aim to maximize return-to-risks like RAROC, RORAC, and RARORAC.

The Euler allocation principle allocates capital according to the risk contributions of each business unit to the overall financial firm. The risk contributions are calculated using partial derivatives of the risk measure with respect to the capital of each business unit (Tasche 2008). This principle is mostly used in the context of RORAC maximization, as seen in Tasche (2008), Buch et al. (2011), and others. The Lagrange method for capital allocation aims to determine the optimal capital that should be allocated in order to maximize the overall NPV, the adjusted net present value (APV), the economic value added (EVA), etc. These are subject to several optimality conditions (e.g., marginal NPV equals marginal cost of risk capital and default risk, cost of risk capital offset by the benefit of reduction in default risk, etc.) and are sometimes linked with managers' utility maximization (Stoughton and Zechner 2007; Erel et al. 2015). The capital-loss distance minimization method allocates capital to each business unit such that the allocated capital and the loss are sufficiently close to each other (Dhaene et al. 2012). Baule (2014) develops a risk capital allocation process using an internal sequential auction, which reflects the managers' diverging preferences in terms of empire building while using the marginal risk contribution as a basis for the allocation decision.

90.3 Internal Capital Budgeting of Financial Firms

For firms in a complete market, capital budgeting disregards frictions such as taxes, bankruptcy costs, agency costs, and information asymmetry. Modigliani and Miller's (1958) theory proves that the firm value is unaffected by the type of its financing sources (i.e., equity or debt) in such a complete market. However, in an incomplete

market, capital budgeting accompanies those frictions, while the choice of financing source affects the firm value. In the case of financial firms, the risk of investments is particularly emphasized, which requires risk-adjusted capital budgeting approaches and financing. In an incomplete market, financial firms typically face three types of friction that are more risk-stressed than those described in traditional finance theories (Stoughton and Zechner 2007), namely:

1. Debt is favored in comparison to equity up to a certain limit based on the risk of the underlying asset.
2. Equity capital must be raised before the investment information becomes available.
3. Regulators associate their risk-taking ability with the equity capital.

Consequently, a large portion of financial firms' assets may be more difficult for outside investors to value with higher uncertainty (or risk) compared to nonfinancial firms due to its intermediary role in the financial markets (Houston et al. 1997). This can create an additional layer of agency problems with the need for contracts and institutions to "monitor the monitor" (Diamond 1984).

Risk capital, which is also referred to as economic capital, represents the smallest amount of equity capital required to cover potential losses (Buch et al. 2011; Merton and Perold 1993; Perold 2005). It is crucial for financial firms to back up their obligations to the liability holders, as well as reducing debt overhang, risk shifting, and bankruptcy problems. However, attempts to meet such obligations are costly and limited, since they impose high monitoring costs on shareholders, while the return is subject to corporate tax (Erel et al. 2015). Thus, financial firms must manage their capital structure so as to source risk capital in the most efficient manner by considering how much risk capital to maintain and how much contingent capital to arrange (Perold 2005) in order to manage the frictions in an incomplete market. This issue is particularly important because regulators now impose stricter capital requirements on financial firms, with risk capital being the principal focus of such requirements. In this sense, efficient risk capital management on the part of financial firms represents the core solution to the problem of capital budgeting. This research study investigates the issue of risk capital budgeting through internal and external financing in financial firms, with a particular focus on banks.

Optimal risk capital allocation from internal financing processes is one of the key approaches to this problem. Risk capital must be allocated for financial firms to evaluate performance, make investment decisions, price products and services, and set compensation (Erel et al. 2015). Return-to-risk optimization from risk capital allocation protects and/or enhances shareholder value (Zaik et al. 1996). It redistributes the total (frictional or opportunity) cost of holding capital across various business lines in order to transfer the cost back to depositors or policyholders in the form of charges. Additionally, the allocation of expenses across business lines is necessary for financial reporting purposes (Dhaene et al. 2012). Capital supportive programs of the government may enhance the resource allocation within the internal structure of banks. For instance, Capital Purchase Program (CPP), the first and most

prominent initiative under Troubled Asset Relief Program (TARP) in the USA, has proven to be highly valuable when the banking system is under stress since the program participating banks have transferred more capital to their subsidiaries than the nonparticipating banks (Mukherjee and Pana 2018).

Previous studies provide a number of theories concerning the risk capital allocation of financial firms in terms of performance measurement and return-to-risk optimization (Buch et al. 2011). More specifically, there exist several strands of literature in this area, including the mathematical finance, insurance-linked, and financial economics contexts (Buch et al. 2011). However, only a limited number of practical solutions accompanied by empirical guidance on optimal internal risk capital allocation can be found in the literature including Kang and Poshakwale (2019). The need for a practical solution becomes increasingly important as financial firms strive to maximize their profitability and shareholder value under conditions of stricter financial regulations and market frictions.

90.3.1 Performance Measures for Financial Firms

There are three major types of performance measurements for financial firms – risk-adjusted return on capital (RAROC), return on risk-adjusted capital (RORAC), and risk-adjusted return on risk-adjusted capital (RARORAC). These are different names to refer to adding interest rate to the economic capital or subtracting the product of the capital costs and the economic capital (Buch et al. 2011). These concepts can be confusing, as clear from the fact that even in case of the identical return-to-risk concept, Bielefeld and Maifarth (2005) referred to it as RORAC while Matten (2000) used the term RAROC (Lampenius 2012). Although this study is based on Buch et al.'s (2011) RORAC maximization model, it is also valid for the three types of return-to-risk measures mentioned in that study. The formulae for each of these three return-to-risk measures are as follows (Zaik et al. 1996; Saita 2007; Lampenius 2012; QFinance 2017):

$$\begin{aligned} \text{RAROC} &= \frac{\text{Risk} - \text{Adjusted Net Income}}{\text{Economic Capital}} \\ &= \frac{\text{Revenue} - \text{Expenses} - \text{Expected Loss} + \text{Net Interest Margin of Capital Change}}{\text{Economic Capital}} \end{aligned}$$

$$\text{RORAC} = \frac{\text{Expected Return}}{\text{Risk} - \text{Adjusted Economic Capital}} = \frac{\text{Net Income}}{\text{Allocated Economic Capital}}$$

$$\begin{aligned} \text{RARORAC} &= \frac{\text{Risk} - \text{Adjusted Expected Return}}{\text{Risk} - \text{Adjusted Economic Capital}} \\ &= \frac{\text{Economic Value Added (EVA)}}{\text{Allocated Economic Capital}} \end{aligned}$$

The RAROC measure was the first to be used in the late 1970s by a group at Bankers Trust (Zaik et al. 1996). It adjusts the risk for the income in the numerator.

Therefore, it not only considers direct revenues and expenses but also considers expected loss and net interest margin from the change in the capital (e.g., transfer prices from allocations, charges for internal services, and overhead and tax allocations) (Zaik et al. 1996). On the other hand, RORAC adjusts the risk of the capital in the denominator to emphasize the riskiness in the capital instead of adjusting the return in the numerator as in the case of RAROC. Moreover, RORAC is the most suitable return-to-risk measure for the Euler allocation principle, which calculates the credit risk as stated by the Basel committee (Tasche 2008). RARORAC adjusts the risk from both the return and capital in the numerator and denominator, respectively. RARORAC outperforms RORAC in terms of risk-adjusted performance measures (RAPMs) but only minutely (Lampenius 2012). However, RARORAC is far more complex in its computation than RORAC. Thus, of the three, RORAC is relatively the most practical and suitable return-to-risk performance measure for financial firms, which emphasizes the risk components in capital, reflects the financial regulation, and is relatively easier to calculate.

90.3.2 The Regulatory Capital Requirement: Basel

Regulatory capital is the minimum capital requirement for financial institutions as imposed by the regulator. The main emphasis is on the equity capital that is its key element (Basel Committee on Banking Supervision 1988). The Basel Committee on Banking Supervision outlines its composition and calculation, which evolves over time. The committee was established by the G10 countries in late 1974 in response to the collapse of the Bretton Woods system, which led to huge foreign currency losses in 1973. It sets minimum regulation and supervision standards for banks; shares supervisory issues, approaches, and techniques; identifies current or emerging global financial risks, etc. Its members have gone up to 28 in 2014 (Basel Committee on Banking Supervision 2015).

90.3.2.1 Basel I

The Latin American debt crisis in the early 1980s heightened the concerns of the Basel Committee regarding the deteriorating capital ratios of the main international banks with international risks. To strengthen the stability of the international banking system and to remove the source of competitive inequality from the differences in national capital requirements, the *Basel Capital Accord* (1988 Accord) was released in July 1988 (Basel Committee on Banking Supervision 2015). This accord, also known as Basel I, required banks to maintain a minimum capital ratio of the total capital (i.e., Tier 1 capital + Tier 2 capital) to risk-weighted assets of 8% by the end of 1992 (Basel Committee on Banking Supervision 2015). The definitions of Tier 1 and Tier 2 capital were nearly identical until the release of Basel II, and they further changed it in Basel III. Tier 1 capital is composed of paid-up share capital/common stock and disclosed reserves, while Tier 2 capital comprises undisclosed reserves, asset revaluation reserves, general provisions/general loan-loss reserves, hybrid (debt/equity) capital instruments, and subordinated debt (Basel Committee on

Banking Supervision 1988). Later, the Basel Committee amended Basel I by additionally regarding market risk, unlike the existing accord that mainly focused on credit risk. Accordingly, they issued the *Market Risk Amendment to the Capital Accord* (Basel Committee on Banking Supervision 2015), which allowed banks to use internal models to measure their market risk, and they introduced Tier 3 capital consisting of short-term subordinated debt to support market risk (Basel Committee on Banking Supervision 1996). The risk-weighted asset calculation also changed in Basel III. In Basel I, five different risk weights were imposed on different categories of assets, as shown below, to determine the total risk-weighted assets in the firm (Basel Committee on Banking Supervision 1988; De Spiegeleer et al. 2014).

- 0%: Cash and OECD government debt
- 10%: Loans extended to domestic public-sector entities
- 20%: Loans given to banks incorporated in an OECD country
- 50%: Loans fully secured by a residential property
- 100%: Loans to the private sector, non-OECD banks, real-estate investments, etc.

90.3.2.2 Basel II

The Basel Committee continued to improve its accord by developing more risk-sensitive capital requirements to reflect the underlying risks and effectively address the financial innovation that had taken place in recent years (Basel Committee on Banking Supervision 2015). After almost 6 years of preparation, the committee released the *Revised Capital Framework* in June 2004 known as Basel II. The key elements of Basel I were retained in Basel II, such as total capital (i.e., Tier 1 capital + Tier 2 capital) requirement ratio to the risk-weighted assets of 8%. However, the following three pillars comprised the main revised frameworks in Basel II compared to Basel I (Wernz 2014; Basel Committee on Banking Supervision 2015).

1. The minimum capital requirements were developed and expanded upon compared with the standard rules in Basel I
2. Supervisory review of capital adequacy and an internal assessment process in an institution
3. More emphasis on the effective use of financial reports on the risk and capital structure of an institution to encourage sound banking practices

Through these three pillars, the Basel Committee aimed to achieve three purposes. Through the first pillar, the committee showed that it considered operational risk together with credit and market risks to quantify risks more precisely for accurate minimum capital requirement calculations. Through the second pillar, the committee strove to have banks use adequate internal risk assessment measures that can cover pension risk or goodwill risk in addition to the risks covered in the first pillar. Through the third pillar, the committee stressed the importance of reporting the risk and capital structure of banks (Wernz 2014; Basel Committee on Banking Supervision 2015). Moreover, Basel II uses the internal ratings-based (IRB)

approach to calculate the risk-weighted asset through the following formula (Basel Committee on Banking Supervision 2006).

$$\text{Correlation (R)} = 0.12 \times \frac{1 - e^{(-50 \times \text{PD})}}{1 - e^{(-50)}} + 0.24 \times \left[\frac{1 - (1 - e^{(-50 \times \text{PD})})}{1 - e^{(-50)}} \right]$$

$$\text{Maturity adjustment (b)} = (0.11852 - 0.05478 \times \ln(\text{PD}))^2$$

Capital requirement (K)

$$\begin{aligned} &= \left[\text{LGD} \times \text{N} \left[(1 - R)^{-0.5} \times G(\text{PD}) + \left(\left(\frac{R}{(1 - R)} \right)^{0.5} \times G(0.999) \right) \right] - \text{PD} \times \text{LGD} \right] \\ &\times (1 - 1.5 \times b)^{-1} \times (1 + (M - 2.5) \times b) \end{aligned}$$

$$\text{Risk-weighted assets (RWA)} = K \times 12.5 \times \text{EAD}$$

where PD is probability of default, LGD is loss given default, EAD is exposure at default, M is maturity, N(x) is the cumulative distribution function of a random variable according to standard normal distribution, and G(x) is the inverse cumulative distribution function of a random variable according to standard normal distribution (Wernz 2014).

90.3.2.3 Basel III

The Basel Committee still needed to strengthen its framework, which it attempted to do in Basel II. The global financial crisis of 2008 revealed that the banking sector had been mispricing credit and liquidity risks and had excess credit growth, poor governance, poor risk management and inappropriate incentive structures (Basel Committee on Banking Supervision 2015). Accordingly, the committee set out *Basel III: International framework for liquidity risk measurement, standards and monitoring* and *Basel III: A global regulatory framework for more resilient banks and banking systems* in December 2010, which revised and strengthened the three pillars laid out in Basel II. It also contained several innovations (Basel Committee on Banking Supervision 2015), as follows:

1. Capital conversion buffer: An additional layer of common equity to help protect the minimum common equity requirement that limits payouts of earnings when breached.
2. Countercyclical capital buffer: It restricts the participation of banks in system-wide credit booms to reduce their losses in credit busts.
3. Leverage ratio: The minimum amount of loss-absorbing capital relative to all banks' assets and off-balance sheet exposures regardless of risk weighting.
4. Liquidity requirements: (1) The liquidity coverage ratio (LCR) requires sufficient cash to cover the 30-day period of stress; (2) the net stable funding ratio (NSFR) is a longer term liquidity ratio meant to address maturity mismatches.

5. Additional proposals for systemically important banks: Requirements for supplementary capital, augmented contingent capital, and strengthened cross-border supervision and resolution.

The capital definition changes in Basel III are as follows: Tier 1 capital comprises Common Equity Tier 1 capital and Additional Tier 1 capital. The Common Tier 1 capital consists of common shares, stock surplus, retained earnings, other accumulated comprehensive income, and other disclosed reserves (Basel Committee on Banking Supervision 2011). Additional Tier 1 capital is part of the Tier 1 capital not included in the Common Equity Tier 1 capital. The additional Tier 1 capital is perpetual in nature, has the possibility of a coupon deferral at the discretion of the issuer, has coupons that should be paid out of distributable profits, is subordinated to senior and subordinated debt, etc. (De Spiegeleer et al. 2014). However, Tier 3 capital has been eliminated from Basel III. The risk-weighted capital calculation in Basel III is almost similar to Basel II, except that Basel III uses the additional multiplier 1.25 to the correlation parameter for regulated financial institutions with assets greater than or equal to \$100 billion and unregulated financial institutions regardless of their size (Basel Committee on Banking Supervision 2011). Among all the accords, Basel III is the strictest financial regulation used these days, which makes it tougher for financial firms to be more speculative in raising profits.

90.3.3 Economic Capital

Financial firms can also estimate their risk with their own internal risk models and derive their minimum required equity capital, known both as “risk capital” and “economic capital.” The risk measures and/or confidence levels of risk capital are different from that of regulatory capital. Financial firms are increasingly using economic capital, with rapid advances in risk quantification methodologies, the rising complexity and sophistication of financial portfolios and supervisory expectations to develop internal processes to assess capital adequacy beyond the regulatory capital adequacy guidelines that may not fully reflect the underlying risks. Economic capital is highly dependent on several factors – risk measures, confidence level, time horizon, risk aggregation method, etc. (Basel Committee on Banking Supervision 2009). These have been explained in further detail below.

First, a financial firm can choose between various risk measures, such as standard deviation, value-at-risk (VaR), expected shortfall (ES), and spectral or distorted risk measures, which can affect its economic capital calculation. There are several criteria for evaluating these risk measures – intuitive, stable, easy to compute, easy to understand, coherent, simple, and meaningful risk decomposition (Basel Committee on Banking Supervision 2009). Of these risk measures, VaR and ES are the most widely used, while each of them has its own advantages and disadvantages. VaR is easier to understand than ES but does not satisfy the coherence criteria. ES is coherent but relatively less easy to interpret than VaR, and its link with desired target credit is not clear (Basel Committee on Banking

Supervision 2009). The Bank for International Settlement (BIS) proposed moving its standardized risk measure from VaR to ES in 2012 after identifying a number of deficiencies with VaR (Basel Committee on Banking Supervision 2013). The spectral or distorted risk measure is effective in assigning different weights to the quantiles of loss distribution, which makes it more powerful than other risk measures. However, it is highly dependent on the underlying loss distribution, and its computation method is complicated to use and understand (Basel Committee on Banking Supervision 2009, 2013).

Second, the economic capital calculation depends on the confidence level it uses. The confidence level is the amount of capital buffer for bankruptcy protection. Thus, it is highly related with the target credit rating. The capital buffer can erode when the available capital resources fail to exceed the economic capital. However, the link between the confidence level and target rating is difficult to establish as the relationship between the ratings and the probability of default (PD) can change due to several factors like the rating agency scope adopted, statistical noise, etc. Therefore, financial firms adopt different ranges of confidence levels, instead of just one, in order to link with target ratings for different purposes. For instance, they use high confidence levels when their main objective is to minimize the risk of bankruptcy, and they use low confidence levels when the goal is to increase profits (Basel Committee on Banking Supervision 2009).

Third, the economic capital is also affected by the time horizon it implements. The choice of horizon depends on several factors, such as asset liquidity, the need for risk management, the market position of the financial firm, etc. In general, the market risk is estimated using very short time horizons, such as days or weeks, while credit risk is measured using relatively long-time horizons like years. The time horizons of regulatory requirements – such as 1 year, 10 days, etc. – also affect the calculation of economic capital (Basel Committee on Banking Supervision 2009).

Fourth, the economic capital depends on the risk aggregation methodology it uses. Some typical risk aggregation methodologies are as follows:

- **Summation:** This method adds individual risk components. It assumes inter-risk correlations are one, and each risk component has an equal weight during summation. It also ignores the benefits of potential diversification.
- **Constant diversification:** This method is the same as the summation method, except that it assumes a fixed level of diversification benefits.
- **Variance-covariance:** This method allows for interactions across the risks. However, these interactions are assumed to be linear and fixed across time. The benefits of diversification depend on the pairwise correlation between the risks.
- **Copulas:** This method combines marginal distributions with the joint probability distributions of risks. It also allows for rich interactions between risks.
- **Full modeling/simulation:** This method simulates the impact of common risk drivers on all risk components through scenario analysis. The distribution of outcomes and economic capital risk measures are calculated based on this. It accounts for all risk interdependencies and effects for the financial firm.

The summation and constant diversification methods are easy to compute and communicate. However, they are relatively crude and not sophisticated enough to consider interactions among different risks. On the other hand, variance-covariance considers the interactions among individual risks. It is also easy and intuitive enough to communicate. Nonetheless, it can be difficult and costly to estimate and may lead to the underestimation of tail losses, since it does not consider the possible non-linearity of risk factors. In contrast, the copulas method is more flexible than variance-covariance since it takes into account the nonlinearity of risk factors, leading to better risk approximation. However, the copulas method requires high-quality inputs, requiring a complete set of individual risk components to produce highly sophisticated parameters. Moreover, it is difficult to calculate and communicate (Basel Committee on Banking Supervision 2009).

90.4 Comparison Between Financial and Nonfinancial Firms

Financial firms are not only highly leveraged and risk sensitive, but their equity capital amount is highly limited to cover the illiquid risks compared to nonfinancial firms. Therefore, they are subject to liquidity risks, which link their capital budgeting process with risk management functions. Financial firms manage their risks either by offsetting them through hedging transactions or by altering their investment policies to manage risks that are not feasible to hedge (Froot and Stein 1998). Their investments are financed largely through debt, while equity in the form of economic capital only supports the risks in the investments. The allocated economic capital shows each segment's contribution to the overall risk (Buch et al. 2011). Financial firms aim to attain the target credit quality measured as credit spread, debt rating, etc., through their optimal capital structure that maximizes their firm values. On the other hand, nonfinancial firms strive to attain an optimal capital structure to achieve their target debt ratio, rather than credit quality, to maximize their value (Erel et al. 2015).

Thus, capital budgeting for financial firms requires a rather different approach than that for nonfinancial firms (Buch et al. 2011). For instance, capital budgeting should aim at maximizing a financial firm's specific performance measures such as RAROC, RORAC, RARORAC, etc., which emphasize the risk component in the return and/or in the capital. These allocation methods for financial firms often accompany various risk measure models as well. These risk measures include VaR, ES, conditional tail expectation (CTE),¹ and so on. Besides, it is crucial for financial firms to consider the financial regulation, including regulatory capital requirement, which evolves over time. The risk-adjusted performance measures such as RORAC and RARORAC heavily regard the risk component in the capital and are helpful in reflecting the regulatory capital requirements. Many internal

¹The CTE measure describes the expected amount of risk that can be experienced, given that a potential risk exceeds a certain threshold value (Cai 2005).

capital budgeting issues (e.g., the agency problem, external capital market dependence, financial shock, capital budgeting methodologies, etc.) in the context of general firms are also related, albeit not entirely, to financial firms' internal capital budgeting. According to the type of firm (as discussed in Sect. 90.2.6), for instance, financial firms are highly leveraged, which may lead to greater incentives to invest earlier than nonfinancial firms. Financial firms are highly leveraged and may react more sensitively to financial shocks (as discussed in Sect. 90.2.7) than nonfinancial firms. Internal capital budgeting for general firms concerns efficient capital or resource allocation or management. However, financial firms are specifically concerned with efficient internal "risk" capital allocation/management as they are credit sensitive, highly leveraged, and possess relatively limited equity capital to cover their illiquid risks compared to nonfinancial firms. Therefore, financial firms employ "risk" capital budgeting to maximize their risk-adjusted performance measures, such as RAROC, RORAC, and RARORAC, which leads to the maximization of shareholder value (as discussed in Sects. 90.2.8 and 90.3.1). On the other hand, nonfinancial firms aim to make investment decisions mostly based on the traditional standard DCF measure (as discussed in Sect. 90.2.8).

90.5 Conclusion

Financial firms are highly leveraged, risk sensitive, and possess limited equity capital to cover their illiquid risks compared to nonfinancial firms. Thus, the internal capital budgeting process and its related issues for financial firms are somewhat different from those for nonfinancial firms. Issues related to internal capital budgeting for firms in general include the agency problem, optimal timing for investment, relative dependence on the external capital market, firm type, firm boundary, market type, financial shock, capital budgeting methodologies, and so on. Some of these issues – such as firm type and capital budgeting methodologies – are different for financial firms and nonfinancial firms. Risk-adjusted performance measures including RAROC, RORAC, and RARORAC, regulatory capital and economic capital are background characteristics that differentiate the internal capital budgeting process for financial firms from that for nonfinancial firms.

Therefore, internal capital budgeting for financial firms has several distinct characteristics compared with nonfinancial firms. Financial firms aim to maximize their risk-adjusted performance measures (e.g., RAROC, RORAC, and RARORAC) leading to the maximization of shareholder value, while in case of nonfinancial firms, the internal capital budgeting procedure is primarily affected by the traditional standard DCF method. In particular, RORAC and RARORAC, which emphasize the risk factors in the capital, are quite helpful in reflecting the regulatory capital requirements that financial firms need to meet. Besides, issues related to internal capital budgeting, such as type of firm, financial shock, and capital budgeting methodologies, are different for financial and nonfinancial firms.

Future extensions can be the empirical test of the quantified impact of internal capital budgeting issues on financial firms' performance and shareholder value. It will be interesting to investigate the directions (positive or negative) and magnitudes of these impacts. The challenge would lie in quantifying the nonfinancial effects through appropriate measures to derive practically meaningful results.

References

- Aguerrevere, F.L. 2009. Real options, product market competition, and asset returns. *The Journal of Finance* 64 (2): 957–983.
- Aktas, N., P.C. Andreou, I. Karasamani, and D. Philip. 2019. CEO duality, agency costs, and internal capital allocation efficiency. *British Journal of Management* 30 (2): 473–493.
- Almeida, H., and D. Wolfenzon. 2005. The effect of external finance on the equilibrium allocation of capital. *Journal of Financial Economics* 75 (1): 133–164.
- Arrfelt, M., R. Wiseman, and G. Hult. 2013. Looking backward instead of forward: Aspiration-driven influences on the efficiency of the capital allocation process. *Academy of Management Journal* 56: 1081–1103.
- Arrfelt, M., R.M. Wiseman, G. McNamara, and G.T.M. Hult. 2015. Examining a key corporate role: The influence of capital allocation competency on business unit performance. *Strategic Management Journal* 36: 1017–1034.
- Bardolet, D., C.R. Fox, and D. Lovallo. 2011. Corporate capital allocation: A behavioral perspective. *Strategic Management Journal* 32: 1465–1483.
- Basel Committee on Banking Supervision. 1988. *International convergence of capital measurement and capital standards (updated to April 1998)*. Basel: Bank for International Settlement.
- . 1996. *Amendment to the capital accord to incorporate market risks*. Basel: Bank for International Settlement.
- . 2006. *Basel II: International convergence of capital measurement and capital standards: A revised framework—Comprehensive version*. Basel: Bank for International Settlement.
- . 2009. *Strengthening the resilience of the banking sector—Consultative document*. Basel: Bank for International Settlement.
- . 2011. *Basel III: A global regulatory framework for more resilient banks and banking systems—Revised version June 2011*. Basel: Bank for International Settlement.
- . 2013. *Fundamental review of the trading book: A revised market risk framework*. Basel: Bank for International Settlement.
- . 2015. *A brief history of the Basel committee*. Basel: Bank for International Settlement.
- Baule, R. 2014. Allocation of risk capital on an internal market. *European Journal of Operational Research* 234: 186–196.
- Bernardo, A.E., J. Luo, and J.J.D. Wang. 2006. A theory of socialistic internal capital markets. *Journal of Financial Economics* 80 (3): 485–509.
- Bielefeld, F., and M. Maifarth. 2005. Traditionelle und moderne Kennzahlen der Gesamtbanksteuerung. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung, Sonderheft* 52: 145–181.
- Bower, J.L., and C.G. Gilbert. 2005. *Linking resource allocation to strategy*. Oxford: Oxford University Press.
- Buch, A., G. Dorfleitner, and M. Wimmer. 2011. Risk capital allocation for RORAC optimization. *Journal of Banking and Finance* 35 (11): 3001–3009.
- Busenbark, J.R., R.M. Wiseman, M. Arrfelt, and H.-S. Woo. 2017. A review of the internal capital allocation literature: Piecing together the capital allocation puzzle. *Journal of Management* 43 (8): 2430–2455.

- Cai, J. 2005. Conditional tail expectations for multivariate phase type distributions. *Journal of Applied Probability* 42 (3): 810–825.
- Coase, R.H. 1937. The nature of the firm. *Economica* 4: 386–405.
- Cremers, K.M., R. Huang, and Z. Sautner. 2011. Internal capital markets and corporate politics in a banking group. *Review of Financial Studies* 24: 358–401.
- Datta, S., R. D'Mello, and M. Iskandar-Datta. 2009. Executive compensation and internal capital market efficiency. *Journal of Financial Intermediation* 18 (2): 242–258.
- De Spiegeleer, J., W. Schoutens, and C. Van Hulle. 2014. *The handbook of hybrid securities. Convertible bonds, CoCo bonds, and bail-in*. Wiley.
- Dhaene, J., A. Tsanakas, E.A. Valdez, and S. Vanduffel. 2012. Optimal capital allocation principles. *Journal of Risk and Insurance* 79 (1): 1–28.
- Diamond, D.W. 1984. Financial intermediation and delegated monitoring. *Review of Economic Studies* 51: 393–414.
- Duchin, R., and D. Sosyura. 2013. Divisional managers and internal capital markets. *The Journal of Finance* 68 (2): 387–429.
- Erel, I., S.C. Myers, and J.A. Read Jr. 2015. A theory of risk capital. *Journal of Financial Economics* 118 (3): 620–635.
- Fernández, A.I., F. González, and N. Suárez. 2013. The real effect of banking crises: Finance or asset allocation effects? Some international evidence. *Journal of Banking and Finance* 37 (7): 2419–2433.
- Froot, K.A., and J.C. Stein. 1998. Risk management, capital budgeting, and capital structure policy for financial institutions: An integrated approach. *Journal of Financial Economics* 47 (1): 55–82.
- Gertner, R.H., D.S. Scharfstein, and J.C. Stein. 1994. Internal versus external capital markets. *The Quarterly Journal of Economics* 109 (4): 1211–1230.
- Glaser, M., F. Lopez-de-Silanes, and Z. Sautner. 2013. Opening the black box: Internal capital markets and managerial power. *The Journal of Finance* 68 (4): 1577–1631.
- Graham, R., C.R. Harvey, and M. Puri. 2015. Capital allocation and delegation of decision-making authority within firms. *Journal of Financial Economics* 115 (3): 449–470.
- Grenadier, S.R., and A. Malenko. 2010. A Bayesian approach to real options: The case of distinguishing between temporary and permanent shocks. *The Journal of Finance* 65 (5): 1949–1986.
- . 2011. Real options signaling games with applications to corporate finance. *Review of Financial Studies* 24 (12): 3993–4036.
- Grenadier, S.R., A. Malenko, and I.A. Strebulaev. 2014. Investment busts, reputation and the temptation to blend in with the crowd. *Journal of Financial Economics* 111 (1): 137–157.
- Hoang, D., and M.E. Ruckes. 2015. Informed headquarters and socialistic internal capital markets. *Review of Finance* 19 (3): 1105–1141.
- Hoechle, D., M. Schmid, I. Walter, and D. Yermack. 2012. How much of the diversification discount can be explained by poor corporate governance? *Journal of Financial Economics* 103 (1): 41–60.
- Houston, J., C. James, and D. Marcus. 1997. Capital market frictions and the role of internal capital markets in banking. *Journal of Financial Economics* 46: 135–164.
- Huang, H.H., H. Huang, and P.T. Shih. 2012. Real options and earnings-based bonus compensation. *Journal of Banking and Finance* 36 (8): 2389–2402.
- Hund, J., D. Monk, and S. Tice. 2010. Uncertainty about average profitability and the diversification discount. *Journal of Financial Economics* 96 (3): 463–484.
- Kang, W.-Y., and S. Poshakwale. 2019. A new approach to optimal capital allocation for RORAC maximization in banks. *Journal of Banking and Finance* 106: 153–165.
- Lamont, O. 1997. Cash flow and investment: Evidence from internal capital markets. *The Journal of Finance* 52 (1): 83–109.
- Lampenius, N. 2012. Are real investment decisions based on risk-adjusted performance measures consistent with maximizing shareholder value? *Journal of Risk* 15 (3): 77–101.

- Lang, L.H., and R.M. Stulz. 1994. Tobin's Q, corporate diversification and firm performance. *Journal of Political Economy* 102: 1248–1280.
- Liebeskind, J.P. 2000. Internal capital markets: Benefits, costs, and organizational arrangements. *Organization Science* 11: 58–76.
- Lovullo, D., A.L. Brown, D.J. Teece, and D. Bardolet. 2020. Resource re-allocation capabilities in internal capital markets: The value of overcoming inertia. *Strategic Management Journal* 41 (8): 1365–1380.
- Lyandres, E., and A. Zhdanov. 2010. Accelerated investment effect of risky debt. *Journal of Banking and Finance* 34 (11): 2587–2599.
- Maksimovic, V., and G. Phillips. 2002. Do conglomerate firms allocate resources inefficiently across industries? Theory and evidence. *Journal of Finance* 57: 721–767.
- Matsusaka, J.G., and V. Nanda. 2002. Internal capital markets and corporate refocusing. *Journal of Financial Intermediation* 11: 176–211.
- Matten, C. 2000. *Managing Bank capital: Capital allocation and performance measurement*. 2nd ed. Wiley.
- Matvos, G., and A. Seru. 2014. Resource allocation within firms and financial market dislocation: Evidence from diversified conglomerates. *Review of Financial Studies* 27 (4): 1143–1189.
- McDonald, R.L. 2006. The role of real options in capital budgeting: Theory and practice. *Journal of Applied Corporate Finance*. 18 (2): 28–39.
- Merton, R.C., and A.F. Perold. 1993. Theory of risk capital in financial firms. *Journal of Applied Corporate Finance* 6 (3): 16–32.
- Meyer, M., P. Milgrom, and J. Roberts. 1992. Organizational prospects, influence costs, and ownership changes. *Journal of Economics and Management Strategy* 1: 10–35.
- Mittendorf, B. 2004. Information revelation, incentives, and the value of a real option. *Management Science* 50 (12): 1638–1645.
- Modigliani, F., and M. Miller. 1958. The cost of capital, corporation finance and the theory of investment. *The American Economic Review* 48 (3): 261–297.
- Mortal, S., and N. Reisel. 2013. Capital allocation by public and private firms. *Journal of Financial and Quantitative Analysis* 48 (01): 77–103.
- Mukherjee, T., and E. Pana. 2018. The distribution of the capital purchase program funds: Evidence from bank internal capital markets. *Financial Markets, Institutions and Instruments* 27 (4): 125–143.
- Ovtchinnikov, A.V., and J.J. McConnell. 2009. Capital market imperfections and the sensitivity of investment to stock prices. *Journal of Financial and Quantitative Analysis* 44 (03): 551–578.
- Ozbas, O., and D.S. Scharfstein. 2010. Evidence on the dark side of internal capital markets. *Review of Financial Studies* 23 (2): 581–599.
- Perold, A.F. 2005. Capital allocation in financial firms. *Journal of Applied Corporate Finance* 17 (3): 110–118.
- QFinance. (2017). *Understanding and calculating RORAC, RAROC, and RARORAC*. Available at: <http://www.financepractitioner.com/operations-management-checklists/understanding-and-calculating-rorac-raroc-and-rarorac> (Accessed: 14 February 2017).
- Rajan, R., H. Servaes, and L. Zingales. 2000. The cost of diversity: The diversification discount and inefficient investment. *The Journal of Finance* 55 (1): 35–80.
- Roper, A.H., and M.E. Ruckes. 2012. Intertemporal capital budgeting. *Journal of Banking and Finance* 36 (9): 2543–2551.
- Saita, F. 2007. *Value at risk and bank capital management: Risk adjusted performance, capital management and capital allocation decision making*. 1st ed. Oxford: Elsevier.
- Scharfstein, D.S., and J.C. Stein. 2000. The dark side of internal capital markets: Divisional rent-seeking and inefficient investment. *The Journal of Finance* 55: 2537–2564.
- Scherer, F. 1999. *New perspectives on economic growth and technological innovation*. Washington, DC: Brookings Institution Press.
- Seru, A. 2014. Firm boundaries matter: Evidence from conglomerates and R&D activity. *Journal of Financial Economics* 111 (2): 381–405.

- Shin, H.H., and R.M. Stulz. 1998. Are internal capital markets efficient? *The Quarterly Journal of Economics* 113 (2): 531–552.
- Stein, J.C. 1997. Internal capital markets and the competition for corporate resources. *The Journal of Finance* 52 (1): 111–133.
- Stoughton, N.M., and J. Zechner. 2007. Optimal capital allocation using RAROC and EVA. *Journal of Financial Intermediation* 16 (3): 312–342.
- Tasche, D. 2008. Capital allocation to business units and sub-portfolios: The Euler principle. In *Pillar II in the new Basel accord: The challenge of economic capital*, ed. A. Resti, 423–453. London: Risk Books.
- Wernz, J. 2014. *Bank management and control: Strategy, capital and risk management*. Springer.
- Weston, J.F. 1970. Mergers and acquisitions in business planning. *Rivista Internazionale di Scienze Economiche e Commerciali*: 309–320.
- Williamson, O.E. 1975. *Markets and hierarchies: Analysis and antitrust implications*. New York: Collier Macmillan Publishers.
- . 1991. Strategizing, economizing, and economic organization. *Strategic Management Journal* 12 (S2): 75–94.
- Wulf, J. 2009. Influence and inefficiency in the internal capital market. *Journal of Economic Behavior and Organization* 72: 305–321.
- Xuan, Y. 2009. Empire-building or bridge-building? Evidence from new CEOs' internal capital allocation decisions. *Review of Financial Studies* 22 (12): 4919–4948.
- Zaik, E., J. Walter, G. Kelling, and C. James. 1996. RAROC at Bank of America: From theory to practice. *Journal of Applied Corporate Finance* 9 (2): 83–93.



Job Security and CEO Compensation

91

James S. Ang and Wei Chen

Abstract

We analyze a sample of 2,914 hiring contracts and show how job security plays a role in pay negotiation. We find a significant impact of job security on CEO pay using both firm- and industry-level measures across US firms. In general, CEO candidates and the board members would trade pay for job security. We find this pay for job security also follows a hierarchy order; when a firm is unable to compensate job risk with a higher pay, the CEO will seek for alternative ways of compensation, including a severance agreement, a shorter grant vesting period, and an incentive plan metric that is easier to achieve. Moreover, job security has no value among risk-takers, and this value increases with CEO's risk aversion. By examining the impact of job security under different market condition, we find that this pay concession disappears when the CEO labor market is more favorable to the CEOs.

Keywords

Job Security · CEO Compensation · Risk Preference · executive compensation · labor market

JEL Classifications

G41 · J24,J33 · J41,J63

J. S. Ang (✉)

Department of Finance, College of Business, Florida State University, Tallahassee, FL, USA
e-mail: jang@business.fsu.edu

W. Chen

School of Business, Fuzhou Institute of Technology, Fuzhou, China

91.1 Introduction

Job security is important to both firms and the executives. Commitment to job security is an essential ingredient to motivate firm innovation (Manso 2011; Bena et al. 2020). Conversely, a less secured job could motivate managers to engage in accounting fraud, such as earning smoothing (Fudenberg and Tirole 1995; DeFond and Park 1997), to protect their positions. Since job insecurity implies uncertainty about future wealth, the consequences could extend beyond the executive's loss of current employment, including reduced future career options (Brickley et al. 1999) and the loss of equity-based compensation (Dahiya and Yermack 2008).

In this chapter, we examine the role of job security in initial pay negotiation between the new CEOs and their boards. More specifically, we examine whether job security has value to the CEO, which should be manifested in negotiating total pay package at hiring. If so, how much, in what way (compensation type), in what circumstance (firm's situations) and to whom (type of managers).

Prior literature studying the relation between job security and worker's income spans mostly in the labor economic literature. Campbell et al. (2007) study a sample of British adults and find that high fears of unemployment are found to be associated with lower levels of wage growth for men and no significant link with wage growth for women. Bertola (1990) finds that wages tend to be lower in high job security countries. Böckerman et al. (2011) find that high excessive worker turnover is positively related to individual wages. Firms can be reluctant to compensate job security risk. Lazear (1990) studies the European data and finds that severance pay requirements can even reduce employment. Dai et al. (2020) study the impacts of life-threatening events, such as terrorist attacks, on CEO compensation and finds that CEOs at terrorist attack-proximate firms prefer cash-based compensation increases over equity-based compensation. All studies suggest the strong evidence that there is a tradeoff between pay and job security. Although a substantial empirical literature explores the relationship between executive compensation and risk-taking (Coles et al. 2006; Kempf et al. 2009; Low 2009), the study of how job security is linked with CEO compensation has merely been discussed before.

The reasons that such area has received relatively little attention are in part because theoretical predictions concerning the effects of the threat are ambiguous, and empirical measurement of its effect is complicated by the endogeneity of CEO power on own pay, which is related to the firm's governance. Sitting CEOs gain power over the board, and could cause the board to initiate changes in firm's investment and financing policies to increase the CEOs' job security; this is related to the problem of entrenchment due to poor governance. Studies have shown that firm governance weakens as incumbent CEO's tenure increases since the board learns about the CEO's ability and the CEO gains bargaining power and negotiates for greater autonomy (Hermalin and Weisbach 2001). Therefore, these circumstances complicate attempts to study the value of job security to the CEOs. To avoid these complications, the prefer laboratory setting to study the value of job security should involve one in which job security is exogenous, i.e., outside the control of the CEO, and pay is negotiated in arm length. This setting occurs when a

new CEO engages in pay negotiation at hiring. The degree of job security in these firms exists before pay negotiation and is thus exogenous, i.e., whether the firms have a penchant to dismiss CEOs, or whether the industry or firm specific risk of failure is greater.

Peters and Wagner (2014) were the first ones to discuss such topic in the field of corporate finance, and empirically showed the link between CEO dismissal risk and CEO compensation. It is important to differentiate our paper to that of Peters and Wagners. We focus on CEO pay negotiations at hiring, such that CEOs can only use ex-ante events to estimate their job risk. On the other hand, Peters and Wagners study a sample that involves mostly sitting CEOs, and use ex-post events (eg. forced turnovers) to estimates one's dismissal risk. In our paper, we provide extensive measures for job security based on both industry-level and firm-level information prior to the hiring event to simulate how CEO candidates evaluate their job risks. Overall, our design is much closer to a real negotiation process such that one can only attempt to use ex-ante events to forecast future possibilities. Our paper also provides a more comprehensive analysis for CEO compensation including the magnitude of pay, structure of pay, severance agreement, rate of disbursement, and the difficulty of obtaining such payment. Moreover, we provide empirical evidence to show how CEO risk preference could play a role in pay contract negotiation. Using the CEO's type based on career advancement, employment status and expertise, we are able to categorize CEOs into different risk preference groups, and study how risk preference modulates the value of job security. In addition, we enrich the research by adding CEO job market condition and examine its impact on the value of job security.

We study 2,914 CEO's hiring contracts (including 1,349 severance agreements, and 1,143 grant vesting schedules) from Execucomp and Incentive Lab during 2000 to 2013. First, we find that job security has a price. CEOs and the board members would trade pay for job security. Second, we find that such compensation of risk follows a hierarchy order. When firms are unable to grant higher pay, CEOs would seek for alternatives to compensate their job risk; these include a severance agreement, a shorter grant vesting period, and an incentive plan metric that is easier achieve. Third, we find this pay concession disappears among risk-takers (i.e. ambitious CEOs or turnaround experts in financially distressed firms), and this price increases with CEO's risk aversion (i.e. external hires, outside industry hires and previously unemployed CEOs). Lastly, by examining the impact of job security under different labor market condition, we find that this premium diminishes under favorable CEO labor market condition.

Our research compliments that of Peters and Wagner (2014), and is the first empirical study to examine the impact of CEO's risk preference and CEO market condition on the value of job security. This chapter also contributes to the existing literature on determinants of CEO pay (Baber et al. 1996; Cadman et al. 2010, 2013; Olaniyi 2019; Göx and Hemmer 2020) and severance contract (Almazan and Suarez 2003; Rau and Xu 2013; Klein et al. 2021), and provides a new perspective on determinants of CEO pay packages and the optimal design of managerial incentive contracts at US firms.

The paper is organized as follows. In Sect. 91.2, we develop our hypotheses in further details. Section 91.3 describes our methodology and data. In section 91.4 we provide empirical analysis on relationship between job security and CEO compensation, and also on how CEO risk preference and labor market condition play a role in the compensation contract. Section 91.5 concludes.

91.2 Literature Background and Hypothesis Development

CEO faces very real risk of dismissal in a firm. If an organization experiences an extreme change in performance, observers tends to attribute such change to the organization leaders (Meindl et al. 1985). When the performance is improved, CEOs are praised with generous rewards (Wade et al. 2006). However, when firms fail, CEOs are often the ones to blame and are likely to be held accountable via dismissal (Zemba et al. 2006). CEOs could be at risk as soon as the start of their tenure. Studies have pointed out that CEO turnovers are more likely to occur in the first few years of their tenure (Sebora 1996; Allgood and Farrell 2000). Therefore, CEOs are expected to take into consideration the risk of job loss at hiring and to negotiate terms that can compensate this risk.

As mentioned earlier, both boards and CEOs value the importance of job security. Better job security can stimulate firm innovation (Manso 2011; Bena et al. 2020), and firms with less job security would lead CEOs to take counterproductive actions to hold on to their jobs, e.g., earning smoothing (Fudenberg and Tirole 1995; DeFond and Park 1997). From the CEO's stand point, dismissal from current employment might lead to reduced future career options (Brickley et al. 1999) and loss of equity-based compensation (Dahiya and Yermack 2008). The value of job security may, therefore, be priced in the managerial labor market. When job security is absent, CEOs may demand greater pay to compensate for the increasing risk of dismissal. On the other hand, the board of directors would try to negotiate a lesser pay package with these CEOs. The result of the pay negotiation is the prediction that there is a tradeoff between job security and pay.

H1: Job security has a price; CEO candidates and the board members would trade pay for job security; Or, more secured jobs pay less.

However, in certain cases when the firms do not have sufficient financial resources to grant greater total pay (i.e. financially distressed firms), CEOs would seek for a more foreseeable, stable, and shorter term pay structure as alternatives; that is higher percentage in cash pay, higher likelihood of signing a severance pay, shorter vesting period, and an incentive plan that is easier to achieve.

H1 Corollary: In case when firms are financially stressed, CEO candidates would demand a more secured pay structure to compensation for low job security.

Matching a person to the job based on ability and characteristics such as risk taking/ aversion is crucial for the success of many organizations and is often a device to retain employees. A substantial empirical literatures argue and find evidence consistent with the matching hypothesis, whereby boards of directors seek to

match the “right” executives to their firm’s unique characteristics (Finkelstein and Hambrick 1997; Gibbons and Murphy 1992). The implication is that good matches are more productive than bad matches. These executives may also self-select into their desired firms. In terms of their compensation package, Lazear (1998) and Abdel-khalik (2003) find that CEOs self-select into compensation schemes that are consistent with their risk preferences. It is important to note that the extent the CEO is concerned with job security may also lead to self-selection into job matches. Firms with a need for either a stable management team or commit to a business strategy may choose to hire a CEO who values job security. On the other hand, Risk-averse (risk-seeking) CEOs may self-select to serve in a company that provides high (low) job security.

Consistent with risk averse employees’ preference for job security, previous literatures find a positive relationship between incentives pay and job risk (Demsetz and Lehn 1985; Core et al. 1999; Oyer 2004; Kini et al. 2020). They show that employees require to be compensated for employment in riskier jobs. Less secured jobs could truncate lifetime incomes if ousted CEO faced diminished prospects for a similar CEO job, or a top managerial position. Dismissals often leave these CEOs the stigma of a “damaged goods.” Ward et al. (1995) study the subsequent career outcomes for CEOs who are fired during 1988 to 1992 and find that 39 of the 60 former CEOs have not resumed an executive career after being dismissed.

More secured CEO jobs provide longer duration and more certain lifetime incomes, which are the employment conditions that the more risk-averse CEOs prefer. Everything else equals, they would demand greater pay in case they should be offered a high job risk position. Less risk-averse CEOs, on the other hand, are more willing to take these higher risk jobs and thus, more likely to be hired into such jobs as they demand less premium to compensate for the risk of job loss.

H2: Job sorting: Price of job security increases with CEO’s risk aversion. Conversely, less risk averse CEOs demand lower job security related pay.

In a more extreme case, when the CEOs are risk seeking, job security should lose its value.

H2 Corollary: Job Security Has No Value Among the Risk-Seeking CEOs

CEO candidates’ availability of job alternatives could dramatically influence CEO’s concern for job security in a particular position. More and better job alternatives reduce their perceived dependence on the boards (Kelley et al. 1967; Bacharach and Lawler 1981; Mannix et al. 1989; Pinkley et al. 1994). When the availability of alternatives is limited, the difficulty of finding another position in case of dismissal is greater. To a CEO job candidate, the availability of job alternative depends both on the number of available CEO job vacancies and the number of similar CEO candidates looking for similar jobs, we examine the effect of supply and demand in the CEO job market. If number of job openings is limited relative to the number of qualified executives for the CEO jobs, i.e., job market is unfavorable to CEO job seekers, hiring firms can afford to offer less generous terms and job candidates are more eager to accept. Accordingly, boards with more secured jobs are more likely to demand CEO candidates to make concession in pay in exchange for greater job

security. Similarly, CEOs in this market condition would not be able to demand a higher premium to compensate for riskier jobs.

H3: Price of job security increases (decreases) when the labor market condition is favorable (unfavorable) to CEO job seekers.

91.3 Methodology and Data

Our sample consists of 2,914 CEO hiring events in the 2000–2013 period drawn from the Execucomp¹; 2,895 of which include the dollar amount of total annual pay, and 1,349 of which include a severance agreement.² To reduce a tenure-related bias in observed pay, we limit our sample to new hires because sitting CEO could gain power over the board during their tenure (Hermalin and Weishbach 2001), and use their control of information and board appointment to entrench themselves, resulting in having greater influence on their own pay and job security. To concentrate on pay of new hires, we can avoid this endogenous influence on pay. Moreover, we use Incentive Lab database to incorporate 1,143 detailed incentive plans at hiring, including the metrics and vesting schedules of the incentive grants. The accounting data are from the Compustat database.

The empirical analyses are carried out in three stages. The first stage verifies the validity of our choice of job security variables. Since we focus only on hiring contracts, the CEO candidates can only use past information of firm to evaluate the security of their potential jobs. Therefore, the proxy for job security should be available from the firm or industry before the arrival of the new CEO. Moreover, for a job security proxy to be valid, it should be highly correlated with the new hired CEO's tenure and the chance of dismissal. The second stage explores how job security plays a role in pay negotiation. More precisely, we investigate whether the new hires can, in turn, use lack of job security to negotiate higher pay with the boards. In the cases when the firms do not have sufficient financial resources to provide greater pay, CEOs could seek for an alternate pay structure to compensate for lower job security; that is, in terms of more cash pay, more likelihood of signing a severance agreement, shorter vesting period and an incentive plan that is easier to achieve. In the third stage, we incorporate CEO risk preference and CEO labor market condition into the study, and discuss how they would modulate the value of job security.

91.3.1 Job Security Measures

We categorize our measures into firm-level and industry-level. The first firm-level measure we use is the average job tenure of the firm's previous CEOs before the appointment of the new CEO. A job with long expected tenure would be a strong

¹To avoid complications, we exclude firms with two CEOs at the same time.

²Execucomp database includes severance information starting from 2006.

indicator of high degree of job security. We argue that this measure captures the overall firm-level risk of job security, and could expect to be highly correlated with the future tenure of the incoming CEO. The job tenure information is gathered from Execucomp. We calculate the average job tenure of a firm by summing up the total years of CEO tenures and then divide by the number of CEOs.

$$\text{Firm_}JS_{1,i,j} = \sum_{l=1}^N \frac{\text{Tenure}_l}{N_{i,j}} \quad (1)$$

where $\text{Firm_}JS_{1,i,j}$ is the level of job security measure by the average past CEO tenure of Firm j in year i ; Tenure_l is CEO l 's tenure; and N is firm's total number of past CEOs.

One might argue that job tenure can be time specific since information about tenure are less during the earlier years in the database. This problem is alleviated because our sample focus on later years alone, which starts at 2000 whereas the starting year for Execucomp is 1993. The fact that newly listed firms have less tenure information can also indicate greater uncertainty. This set up is also in line with our job security definition.

The second firm-level variable is based on outgoing CEO's reason for departure. If the outgoing CEO leaves the firm near the customary retiring age, the chance of him successfully ending his CEO career would be high, and thus generate a positive signal of a secured job to the next CEO. We set $\text{Firm_}JS_{2,i,j}$ equal to 1 when the outgoing CEO left the firm at an age of 60 or greater, and 0 otherwise.

The third firm-level variable we use is an indicator for financially distressed firms. Studies have shown that the likelihood of dismissal is much greater among the distressed firms (Gilson 1989; Engel et al. 2003; Desai et al. 2006). These firms are identified based on Altman's (1968) Z-score. Altman estimates the financial health of a firm using an overall index Z , where

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5 \quad (2)$$

X_1 is the working capital to total assets ratio, X_2 is the retained earnings to total assets ratio, X_3 is the earnings before interest and tax to total assets, X_4 is the market value of equity to total liabilities ratio, and X_5 is the sales to total assets ratio. For firms achieving an Altman's Z-score less than 2.7 are categorized as financially distressed firms, and set $\text{Firm_}JS_{3,i,j}$ equal to 1.

For industry-level job security, we construct measures based on industry market performance, cash flows and credit ratings. Studies have shown that firm's stock market performance are highly correlated with the likelihood of CEO dismissal (Kaplan and Minton 2012; Jenter and Kanaan 2021). We construct our first industry-level measure accordingly by capturing the stock market risk. Which is defined as the ratio of industry annual stock return average to their standard deviation

$$\text{Ind_}JS_{1,i,k} = \sum_{s=1}^M \frac{\text{stock return}_{s,i,k}}{M_{i,k}} / \sigma(\text{stock_return}_k) \quad (3)$$

where $Ind_JS_{1,i,j}$ is the industry-level job security measure by stock market risk in year i and in industry k ; $stock_return$ is firm's annual stock return; M is total number of firms in the industry. The measure is negatively correlated with the industry risk of stock market and therefore positively correlated industry's job security.

The second measure we follow Nguyen (2007)'s methodology, and use the ratio of the industry's average cash flows to their standard deviation to capture cash flow risk.

$$Ind_JS_{2,i,k} = \sum_{s=1}^M \frac{NCF_{s,i,k}}{M_{i,k}} / \sigma(NCF_k) \quad (4)$$

Where NCF is firm's net operating cash flow. The measure is negatively correlated with the industry risk of cash flow and therefore positively correlated industry's job security.

For our third measure, we use S&P long-term issuer credit ratings from Compustat. Compustat maps the official rating categories into integer values from 2 (AAA) to 23 (CC). We further scale these numbers such that a unit increase in our variable corresponds to the difference between an AAA and a BBB rating (the scaling factor is 1/9). Finally, we compute the industry-level job security by computing the average industry credit ratings. The higher $Ind_JS_{3,i,j}$ is the riskier industry, and thus lower industry job security.

$$Ind_JS_{3,i,k} = \sum_{s=1}^M \frac{\text{Credit rating}_{s,i,k}}{M_{i,k}} \quad (5)$$

91.3.2 Compensation Measures

We use the total annual compensation ($tdc1$ scaled to the \$ millions) in Execucomp to measure the total pay described on the hiring contract. This variable consists of the following components: base salary, annual bonus, stock awards, option awards, long term incentive plan, other benefits, and other components.

We use the ratio of annual cash pay to total annual compensation to measure cash payout. Cash pay is calculated as the sum of total annual cash awards plus base salary. In addition to these pay components, we also study the payout when the executive leaves employment at the firm. This is called the severance package which is instituted to help protect the newly unemployed. In general, there are two types of severance agreement. The first type is the severance agreement upon involuntary termination which describes the payment for the named executive officer who are involuntarily terminated but not for cause, e.g., reason other than deliberate and serious disloyal or dishonest conduct. The second type of severance protects the CEOs in case of change-in-control. This agreement describes the payments that are triggered by the occurrence of an acquisition. Severance is a measurement for CEO's protection against dismissal. Therefore, we would expect CEOs to be more likely to demand for a severance agreement when the firm is unable to grant a higher total pay

for job risk. In Execucomp, severance data are available for the years after and including 2006. We define a dummy variable, *severance*, equal to 1 for CEOs who sign severance agreements (i.e., the dollar amount of severance data is non-zero and non-missing).

Firms often break their incentive awards into several vesting periods with amount and schedule based on either year of employment or firm performance. These mechanisms are used to motivate CEO for better performance. However, when the CEO is dismissed, it would also lead to the forfeiture of these unvested compensation. Therefore, a shorter vesting period would provide a greater expected payoff (or greater certainty equivalent) to the employees since it consists of payments that are either guaranteed or foreseeable in the near future. We take advantage of detailed grant vesting schedules provided by the Incentive Lab database, and calculated the weighted average vesting period using a modified method of Gopalan et al. (2014).

$$\text{Average_vesting_period} = \frac{\sum_{i=1}^{n_c} \text{Cash}_i \times t_i + \sum_{j=1}^{n_s} \text{Restricted_stock}_j \times t_j + \sum_{k=1}^{n_o} \text{Option}_k \times t_k}{\sum_{i=1}^{n_c} \text{Cash}_i + \sum_{j=1}^{n_s} \text{Restricted_stock}_j + \sum_{k=1}^{n_o} \text{Option}_k} \quad (6)$$

Where the subscript i denotes a cash grant, the subscript j denotes a restricted stock grant, and subscript k denotes an option grant. Cash is the dollar value of cash grant i with corresponding vesting period t_i . Restricted_stock $_i$ is the dollar value of restricted stock grant i with corresponding vesting period t_i in months. Option $_k$ is the Black-Scholes value of option grant k with the corresponding vesting period t_k in months. In our equation we do not include salary and bonus as the one in Gopalan et al. (2014) because we solely study the vesting schedule of the grants.

Lastly, we measure the ease of incentive plan by considering metrics of the plan. In general, incentive plans are paid out based on either time or performance. A time-based incentive plan is designed to encourage CEOs to work longer for the firm, whereas a performance-based incentive plan is designed to motivate CEOs to achieve certain goal performance. When the high job risk is present, an early dismissal would certainly lead to the forfeiture of unvested time-based compensation. However, under performance-based incentive plan, there is still a chance for CEO to receive the compensation (in the worst case scenario, CEO can manipulate accounting performance). Therefore, time-based incentive plan should be avoided. We calculate time-based portion of the incentive grants by dividing total amount of time-based incentive plan granted by total amount of incentive plan granted.

91.3.3 Risk Preference Measures

Special circumstances may cause the relationship between job security and CEO pay to deviate from the standard prediction above. In the third stage, we point out these

special circumstance. We use information of new hires' previous employment, career advancement, and financial data from Execucomp and Compustat to construct our risk preference measures.

When CEOs possess certain expertise that the firms require, they would have less desire for job security for two reasons. First, the firm would value the expertise of the CEO and should, therefore, be more tolerate of the CEO occasion below par performance which in turn decreases the likelihood of early dismissal. Second, this type of CEOs will likely to be more self-assured that their specialties would cause an impact on the firm, and thus they are more confident of the possible outcomes; this over-confidence which leads them to give less weight to job security. A good example would be a turnaround expert in a financially distressed firm. When a firm desperately needs to change its financial situation, its board will likely to hire a turnaround expert (Ellis 2012). The boards will be more lenient on the CEO who the possesses such turnaround specialty by providing enough time for him to improve the firm. Meanwhile, due to the previous successful experience, these CEO are confident of the possible impact he could bring, and thus would underweight the value of job security.

The turnaround experts in our sample are defined as the executives who has participated in turning a financially distressed firm into a healthy firm. In other words, these executives have increased a firm's Z-score from lower than 2.7 to greater than 3. The dummy variable, $D(Expert)$, equals to 1 if such CEO is a turnaround expert, and 0 otherwise.

We identify ambitious CEO candidates as our second measure of risk preference. These are defined as the ones who are promoted or hired from a rank lower than the third position (we rank executive by their total compensation). This type of candidates has less concerns for job security than the need to excel (or ambitious). At the same time, because the increase in pay is much higher than their original compensation at a lower position, which would could compensate for less job security.

For the third measure, we identify those CEOs who are extremely risk-averse. Campbell et al. (2007) find that workers' fears of unemployment are increased by their previous unemployment experience. Since having successive dismissals on their resume could adversely reduce their prospect to finding the next job, we expect that CEO candidates who were unemployed for a considerable period of time, before taking on a job, would place greater value on job security. We calculate the gap in years between the current and previous employment from Execucomp, and set the $D(Unemployed)$ variable to 1 if the CEO was out of job for more than 2 consecutive years.

Lastly, we create two specific risk preference variables for external hires. Information asymmetry plays a big part in CEO compensation negotiation. Stakeholders who control information can release only enough to manipulate the outcome that they desire (Pfeffer and Lammerding 1981). In general, CEO candidates can be sourced from inside or outside the firm. Internal candidates have more firm-specific human capital than external candidates (Harris and Helfat 1997), and have the access to the first-hand information about the firm, thus they have the upper hand in evaluating firm's current situation comparing to the external candidates. Therefore,

Table 1 CEO Risk Preference Groups

Risk-seeking	Risk-neutral/Unknown	Risk-averse
Ambitious CEO	Non-ambitious	
Turnaround expert	Non-turnaround expert	
	Employed	Unemployed
	Internally hired	Externally hired
	Hired from within industry	Hired from outside industry

it would cause the external hires to value the publicly available firm-level information even more. Similarly, CEOs who are hired outside of the industry would value publicly available industry-level information more than their counterparts.

Following Parrino (1997), we define an external hire to be an executive who has been appeared for the first time with the firm. All others are classified as external hires. The dummy variable $D(External)$ equals to 1 if such new CEO is an external hire and 0 otherwise. We also use Execucomp to verify CEO’s previous employment. If the CEO was hired from outside industry, we set $D(Outside_Ind)$ to 1 and 0 otherwise.

As shown in Table 1 We separate our CEOs into three categories based on their risk preference, risk-seeking, risk-neutral/unknown and risk-averse.

91.3.4 CEO Labor Market Condition Measures

For CEO labor market condition, we calculate the demand/supply ratio of the entire CEO labor market and the market within the same industry during each year. The demand is calculated as total number of CEO turnovers (new CEO openings) plus total number of incumbent CEOs who are older than 60 in a given year. The supply is calculated as the total number of dismissed CEOs who are 55 or under in the previous year plus the total number of executive who are 55 or under in number 2 and 3 position in a given year.

Table 2 summarizes our key variables (see Appendix A for variable definitions). On average, the total annual pay on hiring year is about \$5.28 million. Of which, 37.3% consists of cash pay. Since the availability of severance data on Execucomp from 2006, 83.3% of the new hires sign a severance agreement. Using the Incentive Lab sample, we find that the weighted average grant vesting period is around 38.51 months. For all the CEOs in our sample, 32.9% are ambitious, 4.5% were unemployed before the appointment, 78% are internally promoted CEOs, and 13.4% are turnaround experts.³ The average age at hiring is 52.64. One other thing to notice

³These categories are not mutually exclusive.

Table 2 Summary Statistics. The sample consists of all CEO who are hired or promoted during 2000–2013 in Execucomp firms. The data for severance agreement are available from 2006–2013 only. The information on vesting period are taken from Incentive Labs. All variables except dummies are winsorized at the 1st and 99th percentiles

	Obs	Mean	Std. Dev.	Min	Max
<i>Compensation</i>					
Total pay (\$million)	2895	5.283	6.562	0.135	38.247
Cash/Total pay	2891	0.373	0.279	0.000	1.000
Severance	1620	0.833	0.373	0.000	1.000
Average vesting period	1116	38.513	14.098	0.170	88.730
Time based grant ratio	1143	0.726	0.321	0.000	1.000
<i>Job security</i>					
Firm_JS ₁	2895	8.709	6.306	1.000	36.000
Firm_JS ₂	2854	0.415	0.493	0.000	1.000
Firm_JS ₃	2380	0.441	0.497	0.000	1.000
Ind_JS ₁	2528	0.095	0.583	−2.222	2.440
Ind_JS ₂	2815	0.185	0.346	−0.065	2.163
Ind_JS ₃	1581	1.399	0.194	0.937	1.993
<i>CEO Characteristics</i>					
Ambitious	2914	0.329	0.470	0.000	1.000
Unemployed	2536	0.045	0.206	0.000	1.000
Outsider	2914	0.220	0.414	0.000	1.000
Insider	2914	0.780	0.414	0.000	1.000
Expert	2914	0.134	0.341	0.000	1.000
CEO age	2891	52.644	7.008	30.000	80.000
Outside Industry	2536	0.040	0.197	0.000	1.000
<i>Market condition</i>					
Demand/supply	2914	0.363	0.481	0.000	1.000
Industry demand/supply	2880	0.339	0.473	0.000	1.000
<i>Firm Characteristics</i>					
lnSize	2832	7.681	1.868	3.308	12.532
Sales growth	2818	0.049	0.239	−0.587	1.243
ROA	2829	0.001	0.162	−0.859	0.305
Leverage	2817	0.237	0.210	0.000	0.991
M/B	2769	2.524	3.237	−9.614	20.332
R&D	2822	0.042	0.090	0.000	0.558
Stock volatility	2661	0.031	0.018	0.009	0.102
Industry-adj stock return	2542	−0.101	0.442	−1.231	1.445
Firm age	2612	24.947	19.26	0.038	88.06

is that 44.7% (*Firm_JS₃*) of the firms are financially distressed. Along with the close-to-zero average ROA and negative stock performance in our sample, our results are consistent with prior literature that a CEO turnover occurs when firm performance is poor (Meindl et al. 1985; Zemba et al. 2006).

91.4 Empirical Analysis

91.4.1 Verification of Job Security Measures

In this section, we verify the validity of the two types of job security measures we discussed in the previous section.

Figure 1a and b depict the relationship between our choice of job security measures and future CEO tenures. We separate our sample into high and low job security groups based on the median of each continuous job security variable. For dummy variable

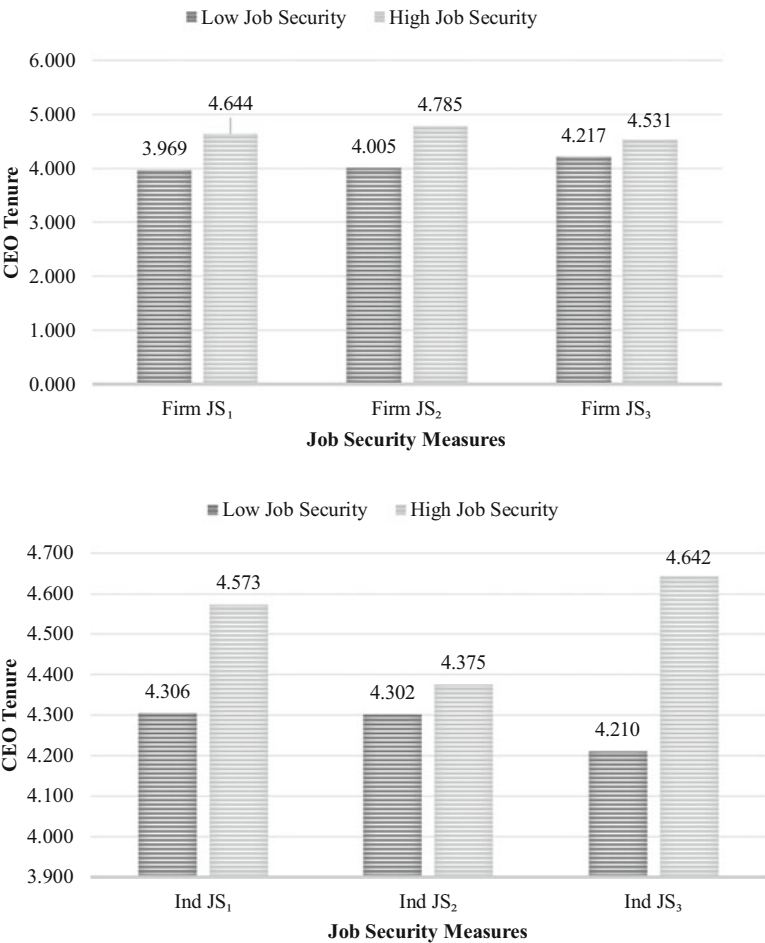


Fig. 1 (a) Average Tenure of Newly Hired CEOs Based on Firm-level Job Security Measures. The sample consists of all CEO who are hired or promoted during 2000–2013 in an Execucomp firm. We separate the level of job security into two groups based on the median of each firm-level job security measure. (b) Average Tenure of Newly Hired CEOs Based on Industry-level Job Security Measures

Firm_JS₃, we categorize as high job security when it equals to 0, and low job security if it equals to 1. We observe difference between the two groups. The new hires in high security jobs have longer tenure than the ones in low security jobs, and the results are significant at 5% level at least across all six measures.

A valid job security measure should also capture CEO's early dismissal risk. Figure 2a and b depicts the relationship between our choice of job security measures and the chance of dismissal during the first 3 years of tenure. The chance of dismissal is calculated as the total number of dismissed CEOs during the first 3 years of tenure divided by the total number of new CEOs at year 0. Again, we find that CEOs are at greater risk of dismissal in low security jobs than its counter-group across all six measures. Thus, these results verify our choice of job secure measures.

91.4.2 Job Security and CEO Compensation

In this section, we examine the relationship between CEO compensation and the two main types of job security measures discussed in the previous section: industry-level measure and firm-level measure. Due the large amount of missing corporate governance data, we consider regression specifications in which control variables are added sequentially. The baseline specification includes only firm characteristics and CEO characteristics, and the second specification includes corporate governance controls is reported in our robustness test section.

In Table 3, we investigate the impact of industry-level job security in column 1 to 3. First, our results indicate that industry-level job security is highly correlated with CEO total pay. Consistent with Hypothesis 1, CEOs are compensated for higher industry job risk; or more secured job pay less. We calibrate the economic significance of these findings. On average, each unit increase in the inverse coefficient of variation (or one-unit decrease in risk) in industry-average stock return and cash flow would lead would lead to a \$984 thousand and \$914 thousand decrease in total pay respectively for newly hired CEOs. Each unit increase in industry-average credit rating (or from an AAA to a BBB, or an BBB to CCC, etc...), new hires' total pay would increase by \$5 million.

In the latter half of Table 3, we investigate the impact of firm-level job security. In column 3 and 4, our results indicate that firm-level job security is highly correlated with CEO total pay. Again, consistent with Hypothesis 1, CEOs are compensated for higher firm-level job risk. On average, each year decrease in average past CEO tenure would lead to a \$24 thousand increase in total pay for newly hired CEOs. New hires' total pay would decrease by \$405 thousand if their predecessors have successfully retired.

In Column 6, we find that firm-level job security measured by financially condition produces a seemingly contrary result; more secured jobs pay more. This situation is further analyzed in the next section.

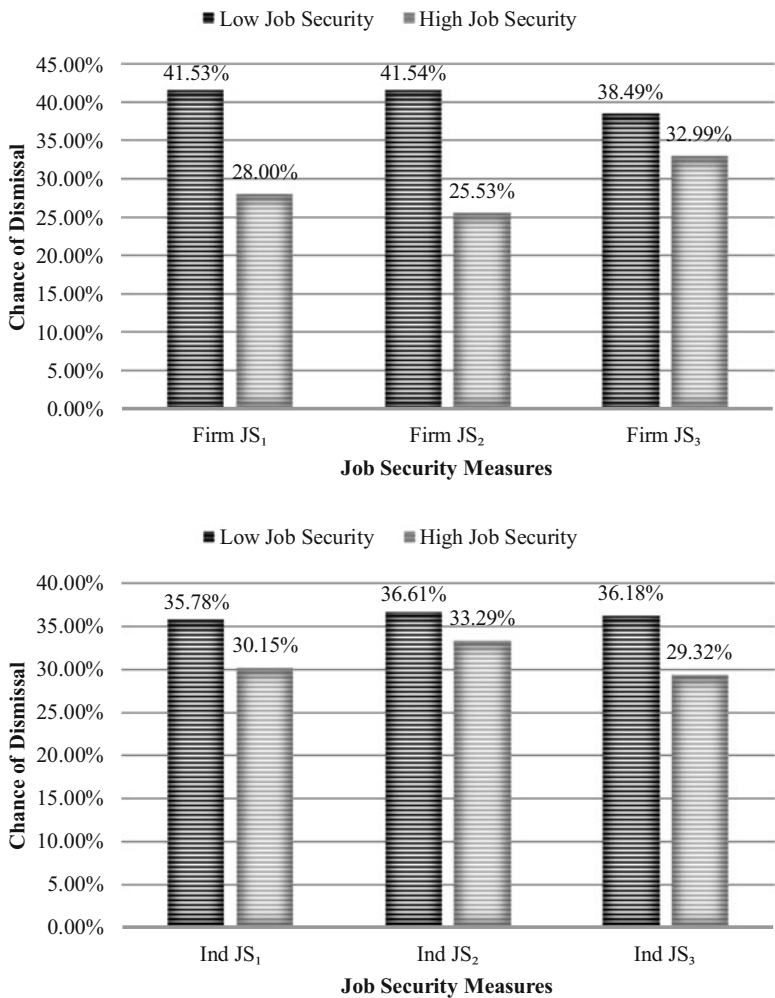


Fig. 2 (a) Rate of Dismissal within the First 3 Years Based on Firm-level Job Security Measures. The sample consists of all CEO who are hired or promoted during 2000–2013 in an Execucomp firm. We separate the level of job security into two groups based on the median of each industry-level job security measure. The chance of dismissal is calculated as number dismissed CEOs during the first 3 years of tenure divided by the total number of CEOs. (b) The sample consists of all CEO who are hired or promoted during 2000–2013 in an Execucomp firm. We separate the level of job security into two groups based on the median of each firm-level job security measure. The chance of dismissal is calculated as number dismissed CEOs during the first 3 years of tenure divided by the total number of CEOs

Table 3 Regression Analysis for Job Security and CEO Pay. Column 1 to 3 report the OLS regression of industry-level job security measure on CEO total pay. *Ind_JS₁* and *Ind_JS₂* are the inverse coefficient of variation in industry stock returns and cash flow, which both are positively correlated with industry job security. *Ind_JS₃* is the industry average credit rating (inversely correlated with industry job security). Column 3 to 6 report the OLS regression of firm-level job security measure on the CEO total pay. Both *Firm_JS₁* and *Firm_JS₂* are positively correlated with firm job security; *Firm_JS₃* is the average tenure of past CEOs and *Firm_JS₂* is an indicator whether the predecessor successfully retired. *Firm_JS₃* is an indicators equal to 1 if the firm is financially distressed. All other independent variables are defined in [Appendix A](#). Robust standard errors are reported in parentheses

Dependent Variable:	Industry-level			Firm-level		
Total pay	(1)	(2)	(3)	(4)	(5)	(6)
Ind_JS ₁	-0.984** (-2.31)					
Ind_JS ₂		-0.914*** (-3.36)				
Ind_JS ₃			5.461*** (5.43)			
Firm_JS ₁				-0.024* (-1.66)		
Firm_JS ₂					-0.405** (-2.43)	
Firm_JS ₃						-0.725*** (-2.92)
Size	1.977*** (21.68)	2.003*** (20.80)	2.495*** (16.28)	2.251*** (22.20)	2.271*** (16.15)	2.272*** (20.55)
Sales growth	1.221* (1.94)	1.066* (1.67)	1.484 (1.41)	1.218** (2.00)	1.249** (2.03)	0.863 (1.36)
ROA	-1.642* (-1.69)	-0.885 (-0.88)	-2.057 (-0.76)	-1.625 (-1.58)	-1.576 (-1.21)	-1.970* (-1.91)
Leverage	-1.786*** (-3.06)	-1.708*** (-2.88)	-2.308** (-2.22)	-1.792*** (-3.12)	-1.642** (-2.12)	-1.804*** (-2.98)

MB	0.225*** (5.19)	0.223*** (5.11)	0.229*** (3.74)	0.193*** (3.68)	0.170*** (4.40)
RD	9.111*** (6.63)	9.092*** (6.63)	14.702*** (3.95)	7.005*** (4.63)	5.880*** (3.64)
Stock return		20.470** (2.09)	-3.221 (-0.19)	17.753* (1.87)	29.099*** (2.88)
volatility		0.537* (1.93)	0.912* (1.87)	0.724*** (3.37)	0.679** (2.52)
Industry-adj.	0.423 (1.55)				
stock return					
D(Insider)	-2.899*** (-8.58)	-2.886*** (-8.52)	-3.630*** (-6.73)	-2.869*** (-8.66)	-3.032*** (-8.83)
CEO age	-0.052*** (-3.49)	-0.050*** (-3.41)	-0.071*** (-2.68)	-0.048*** (-3.25)	-0.047*** (-3.29)
Firm age	-0.003 (-0.49)	0.000 (0.05)	-0.004 (-0.57)	0.000 (0.04)	0.004 (0.63)
Constant	-4.243*** (-3.87)	-5.502*** (-4.48)	-14.521*** (-5.71)	-7.430*** (-5.21)	-8.352*** (-7.17)
Industry FE	2,441	2,441	1,336	2,441	2,117
Year FE	NO	NO	NO	YES	YES
Observations	YES	YES	YES	YES	YES
Adj. R-squared	0.319	0.321	0.288	0.357	0.394

The symbol *** indicates that the p-value is less than 0.01, ** that it is less than 0.05, and * that it is less than 0.1

91.4.3 Job Security and CEO Compensation in Firms with Financial Constraints

The above analyses cover the general cases. However, in a special case when the firms do not have sufficient financial resources to grant greater total pay, we would expect CEOs to seek for alternatives to compensate for low job security.

In Panel A of Table 4, we conduct a univariate analysis for total pay scaled by firm size based on different firm-level job security measures. Indeed, we find that new hires in financially distressed firms receive a lower total pay. This shows distressed firms are unable to grant CEOs with the same pay level as the normal firms. Therefore, CEOs in this type of firms would seek alternative ways to compensate the job risk. And this is exactly what we find in the latter Panels of Table 5. In Panel B, we find that CEOs in financially distressed firms demand 2.395% higher in

Table 4 Univariate Analysis for Financially Distressed Firms. We separate our sample into high and low job security firms based on the median of each firm-level job security measure. The total annual pay, cash pay portion, average vesting period and time-based portion of the grant are compared between two groups

Panel A: Annual Pay (scaled by size)						
	Low Job Security		High Job Security		Difference	T-value
	Obs.	Mean	Obs.	Mean		
Firm_JS ₁	1349	0.433%	1337	0.300%	0.133%	5.5676
Firm_JS ₂	2371	0.369%	442	0.336%	0.034%	1.0642
Firm_JS ₃	909	0.368%	1125	0.440%	-0.072%	-2.7112
Panel B: Cash pay/Annual pay						
	Low Job Security		High Job Security		Difference	T-value
	Obs.	Mean	Obs.	Mean		
Firm_JS ₁	1384	37.211%	1379	37.400%	-0.189%	-0.178
Firm_JS ₂	1663	36.940%	1185	37.500%	-0.560%	-0.530
Firm_JS ₃	1051	37.984%	1331	35.588%	2.395%	2.101
Panel C: Vesting period						
	Low Job Security		High Job Security		Difference	T-value
	Obs.	Mean	Obs.	Mean		
Firm_JS ₁	457	36.724	479	39.888	-3.164	-3.501
Firm_JS ₂	313	38.809	267	39.380	-0.572	-0.502
Firm_JS ₃	408	36.681	525	39.470	-2.789	-3.095
Panel D: Time based grant/Total incentive grant						
	Low Job Security		High Job Security		Difference	T-value
	Obs.	Mean	Obs.	Mean		
Firm_JS ₁	473	71.825%	485	70.851%	0.974%	0.470
Firm_JS ₂	320	71.764%	273	67.850%	3.914%	1.488
Firm_JS ₃	408	69.484%	525	75.605%	-6.121%	-2.941

Table 5 Regression Analysis for Financially Distressed Firm and CEO Pay. Column 1 reports the result of OLS regression. The dependent variable is CEO cash pay ratio. In Column 2, the dependent variable is an indicator that is one if CEO signs a severance agreement and zero otherwise. Logit regression model is used and the marginal effects of the coefficients are reported. Column 3 and 4 report the results of Tobit regressions; the dependent variable are average vesting period and time based portion of the total grant respectively. The independent variables are defined in [Appendix A](#). Robust standard errors are reported in parentheses

	(1)	(2)	(3)	(4)
Dependent variable:	Cash/Total pay	D (Severance)	Vesting period	Time based grant/Total grant
Firm_JS ₃	0.031** (2.38)	0.034* (1.69)	-1.765** (-3.16)	0.035 (1.57)
Size	-0.057*** (-13.63)	0.012 (1.39)	0.780 (1.68)	-0.027*** (-3.11)
Sales growth	0.036 (1.13)	-0.019 (-0.43)	1.185 (1.15)	-0.056 (-1.16)
ROA	-0.040 (-0.77)	0.074 (1.05)	14.070*** (21.23)	-0.042 (-0.49)
Leverage	-0.051 (-1.40)	0.029 (0.54)	-0.551 (-0.22)	-0.026 (-0.24)
MB	-0.007*** (-4.47)	0.004 (1.16)	-0.046 (-0.78)	-0.001 (-0.53)
RD	-0.437*** (-5.40)	-0.115 (-0.99)	4.170 (0.82)	0.186 (1.40)
Stock return volatility	0.313 (0.58)	-2.524*** (-3.42)	25.857 (0.75)	0.399 (0.39)
Industry-adj. stock return	-0.046*** (-3.19)	-0.009 (-0.39)	1.338 (1.51)	0.043** (2.33)
D(Insider)	0.153*** (11.45)	0.009 (0.46)	-3.194* (-2.76)	-0.035 (-1.27)
CEO age	0.004*** (4.85)	-0.004*** (-2.83)	0.001 (0.03)	-0.001 (-1.43)
Firm age	0.001*** (2.99)	-0.001 (-1.10)	-0.261*** (-5.88)	0.001 (0.63)
Constant	0.534*** (4.56)		50.143*** (12.87)	1.118*** (8.64)
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	2,115	1,103	729	744
Adj. R-squared			0.140	
Pseudo R-squared	0.608	0.188		0.667

The symbol *** indicates that the p-value is less than 0.01, ** that it is less than 0.05, and * that it is less than 0.1

cash pay than those in normal firms. We also include alternative job security measures and show that there is no significant difference in the general case. We make use of the Incentive Lab database, and compare the vesting periods and the metrics of incentive plans between high and low job security groups in Panel C. Again, we confirm that CEOs in distressed firm demand a shorter vesting period. In Panel D we find that CEOs in distressed firm tend to choose a metric that is easier to achieve. Using alternative measures, we do not find a similar result. All these results indicate that compensation for job risk follows a hierarchy; CEO will ask for alternate forms of compensation when the firm cannot grant higher pay.

In Table 5, we conduct a more rigorous analysis using regression, and again confirm our findings in the univariate analysis. In terms of economic significance, when the firm cannot grant a higher pay, CEOs on average demand 3.1% higher in cash pay, 3.4% increase in the likelihood of signing a severance agreement, and shortened average vesting period by 1.765 months in less secured jobs. In column 4, we do not find a significant impact on the choice of metrics of incentive plans. We suspect that is due to the differences in CEO risk preference, which will be discussed in more details in the next section.

91.4.4 Job Security under Different Risk Preferences

We would expect the above findings to hold under normal conditions. However, special circumstances may cause the payment scheme to deviate from the standard prediction above. In this section we study the impact of job security under different CEO risk preference. Table 6, we separate our sample based on how CEOs were promoted to current position. We classify those ambitious CEO are those who were promoted from a lower rank to be labelled as ambitious (or risk-seeking). Corroborate with our prediction in Hypothesis 2, job security has a value only among the non-ambitious CEO and has no impact on the CEO's total pay among ambitious ones. From Table 5 in the previous section, we suspect that CEO's risk preference causes the insignificant correlation between job security and the choice of incentive plan metrics. The findings in column 5 and 6 of Table 6 confirm this suspicion. We find CEOs, in financially distressed firms, choose a metric that is easier to achieve to compensation the job risk only if they are non-ambitious. Again, these findings support Hypothesis 2 such that job security has no value among the risk seeking CEOs.

Next, we examine whether external hires would value job security differently from the internal hires. We first study the firm-level measures in Column 1 and 2. The interaction term of external hire indicator and job security measures demonstrate that external hires value the firm-level security more than the internal hires. On average, external hires require a \$97 thousand higher premium for each year decrease in average past CEO tenure than their counterparts. At the industry-level, we compare the difference between CEOs who are hired from outside industry and within industry in Column 3 and 4. The signs of interaction terms of outside-industry hire indicator and job security measures are same as we expected, and are significant

Table 6 Regression Analysis for CEOs with different Risk Preference. Column 1 to 4 report the OLS regression of firm-level job security on CEO total pay among ambitious and non-ambitious CEOs respectively. Ambitious CEOs are defined as the ones who were promoted or hired from a rank lower than the 3rd position. Column 5 and 6 report the Tobit regression of job security on time based grant ratio (left censored at 0 and right censored at 1). The independent variables are defined in [Appendix A](#). Robust standard errors are reported in parentheses

Dependent Variable	Total pay		Time based grant/Total grant			
	Ambitious (1)	Non-ambitious (2)	Ambitious (3)	Non-ambitious (4)	Ambitious (5)	Non-ambitious (6)
Firm_JS ₁	-0.007 (-0.22)	-0.028* (-2.06)				
Firm_JS ₂			-0.534 (-1.14)	-0.469** (-2.43)		
Firm_JS ₃					0.009 (0.17)	0.073*** (2.65)
Size	2.418*** (11.53)	2.243*** (13.68)	2.443*** (11.57)	2.234*** (15.98)	-0.053*** (-2.65)	-0.014 (-1.37)
Sales growth	0.439 (0.38)	2.053** (2.98)	0.481 (0.42)	2.329*** (2.79)	-0.150 (-1.46)	0.016 (0.19)
ROA	-1.223 (-0.67)	-1.789 (-1.56)	-1.403 (-0.72)	-1.818 (-1.19)	-0.083 (-0.41)	-0.145 (-1.14)
Leverage	-2.403** (-2.09)	-1.436*** (-3.64)	-2.714** (-2.28)	-1.629* (-1.89)	-0.048 (-0.30)	-0.052 (-0.43)
MB	0.210** (2.22)	0.166** (2.86)	0.218** (2.24)	0.166*** (3.29)	0.002 (0.38)	-0.001 (-0.23)
RD	7.083** (2.01)	6.130*** (5.20)	6.699* (1.84)	6.183*** (4.21)	0.376 (1.55)	-0.001 (-0.01)
Stock return volatility	25.471 (1.34)	18.339 (1.31)	21.431 (1.10)	17.034 (1.35)	1.073 (0.42)	0.887 (0.75)
Industry-adj. stock return	-0.013	1.049**	-0.023	1.027***	0.027	0.073*

(continued)

Table 6 (continued)

Dependent Variable	Total pay		Time based grant/Total grant	
	Ambitious	Non-ambitious	Ambitious	Non-ambitious
D(Insider)	(-0.02) -3.000*** (-6.13)	(2.78) -3.378*** (-5.16)	(-0.04) -2.968*** (-5.98)	(3.63) -3.323*** (-6.37)
CEO age	-0.110*** (-4.07)	-0.002 (-0.12)	-0.109*** (-3.93)	-0.003 (-0.18)
Firm age	0.006 (0.36)	-0.002 (-0.20)	0.007 (0.39)	-0.003 (-0.33)
Constant	-4.401 (-1.17)	-10.477*** (-5.24)	-4.449 (-1.14)	-9.904*** (-6.56)
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	745	1,696	734	1,675
Adj. R-squared	0.324	0.385	0.326	0.376
Pseudo R-squared			0.431	0.826

The symbol *** indicates that the p-value is less than 0.01, ** that it is less than 0.05, and * that it is less than 0.1

Table 7 Regression Analysis for Unemployed CEOs. The results of OLS regressions are reported in this table. The dependent variable in Table 7 is CEO total pay. $D(\text{External})$ is an indicator that equals to 1 if the CEO was hired externally. $D(\text{Outside industry})$ is an indicator equal to 1 if the CEO was hired from outside industry. The independent variables are defined in Appendix A. Robust standard errors are reported in parentheses

Dependent variable:	Firm-level		Industry-level		
Total pay	(1)	(2)	(3)	(4)	(5)
D(External) *Firm_JS ₁	−0.097**				
	(−2.13)				
D(External) *Firm_JS ₂		−0.336			
		(−0.51)			
D(Outside industry) *Ind_JS ₁			−1.698		
			(−1.20)		
D(Outside industry) *Ind_JS ₂				−6.398**	
				(−2.01)	
D(Outside industry) *Ind_JS ₃					14.401**
					(2.66)
Firm_JS ₁	0.002				
	(0.10)				
Firm_JS ₂		−0.449**			
		(−2.01)			
Ind_JS ₁			1.600		
			(0.70)		
Ind_JS ₂				0.042	
				(0.02)	
Ind_JS ₃					3.317
					(0.57)
D(External)	3.719***	2.936***			
	(8.01)	(6.94)			
D(Outside Industry)			−0.895	−0.354	−23.037**
			(−1.00)	(−0.32)	(−2.98)
Size	2.128***	2.253***	2.387***	2.361***	2.850***
	(30.48)	(22.09)	(6.36)	(6.61)	(3.14)
Sales growth	1.193**	1.325**	1.103	1.475	−1.228
	(2.41)	(2.16)	(0.37)	(0.51)	(−0.39)
ROA	−1.417	−1.616	1.157	−0.762	0.994
	(−1.59)	(−1.53)	(0.23)	(−0.15)	(0.10)
Leverage	−1.655***	−1.869***	−3.762	−3.570	−4.366
	(−2.89)	(−3.18)	(−1.27)	(−1.23)	(−0.95)
MB	0.217***	0.194***	0.494***	0.534***	0.405**
	(6.44)	(4.64)	(2.95)	(3.33)	(2.73)

(continued)

Table 7 (continued)

Dependent variable:	Firm-level		Industry-level		
Total pay	(1)	(2)	(3)	(4)	(5)
RD	4.925***	6.788***	12.008	10.736	5.122
	(3.25)	(4.41)	(1.39)	(1.21)	(0.48)
Stock return volatility	21.204**	15.758		-47.573	-130.018
	(2.50)	(1.62)		(-1.11)	(-1.37)
Industry-adj. stock return	0.608**	0.687**	-0.831	-1.488	-0.720
	(2.32)	(2.43)	(-0.47)	(-0.82)	(-0.22)
CEO age	-0.050***	-0.046***	-0.382***	-0.384***	-0.458***
	(-3.28)	(-3.04)	(-4.62)	(-4.82)	(-3.84)
Firm age	-0.003	0.000	0.036	0.038	0.042
	(-0.46)	(0.02)	(1.12)	(1.20)	(1.26)
Constant	-8.843***	-10.375***	9.528*	11.744**	10.358
	(-7.96)	(-7.14)	(1.81)	(2.16)	(0.81)
Observations	2,441	2,409	189	189	112
Industry FE	YES	YES	NO	NO	NO
Year FE	YES	YES	YES	YES	YES
Adj. R-squared	0.339	0.356	0.294	0.308	0.260

The symbol *** indicates that the p-value is less than 0.01, ** that it is less than 0.05, and * that it is less than 0.1

at 5% level in Column 4 and 5. These findings confirm that CEOs hired from outside industry value industry-level job security more than those who are hired within. In terms of economic significance, CEOs hired from outside-industry would demand a \$6.39 million higher premium for each unit decrease in the inverse coefficient of variation (or one-unit increase in risk) in industry-average cash flow than their counterparts; for each unit decrease in industry-average credit rating (or from an AAA to a BBB, or an BBB to CCC, etc. . .), outside-industry hires would demand a \$14.40 million higher premium than inside industry hires (\$1.66 million higher premium from an AAA to a AA or AA to A, etc...).

We expect that CEO candidate who were unemployed for a long period of time, before taking on the job, should especially value job security. Consistent with our prediction, the interaction term of job security and unemployment indicator is strongly correlated with total pay in Column 1 of Table 8. On average, CEOs who were unemployed previously require a \$197 thousand higher premium for each year decrease in average past CEO tenure than their counterparts. We also develop a case specifically for financial distressed firms in column 3. The interaction term of turnaround expert indicator and job security measure is negatively correlated with the likelihood of signing a severance agreement. As an economic interpretation, turnaround experts in financially distressed firms are 10.6% less likely to sign a severance agreement as a compensation for job risk than their counterparts.

Table 8 Regression Analysis for Turnover Experts. Column 1 and 2 reports the results of OLS regression. The dependent variables in Column 1 and 2 are CEO total pay. *D(Unemployed)* is an indicator that equals to 1 if the CEO was previously unemployed. The dependent variable in Column 3 is an indicator whether the CEO signs a severance agreement. Logit regression model is used in this column and the marginal effects of the coefficients are reported. The independent variables are defined in [Appendix A](#). Robust standard errors are reported in parentheses

Dependent variable	Total pay		D(Severance)
	(1)	(2)	(3)
D(Unemployed)* Firm_JS ₁	−0.197*** (−2.60)		
D(Unemployed)* Firm_JS ₂		1.538 (0.91)	
D(Expert)*Firm_JS ₃			−0.102** (−2.11)
D(Expert)*Firm_JS ₃			0.094*** (2.76)
D(Unemployed)	−2.930*** (−2.60)	−4.862*** (−5.53)	
Firm_JS ₁	−0.021 (−1.40)		
Firm_JS ₂		−0.424* (−1.90)	
Firm_JS ₃			0.046** (2.17)
Size	2.169*** (20.84)	2.175*** (20.42)	0.010 (1.20)
Sales growth	1.272** (2.01)	1.302** (2.01)	−0.017 (−0.42)
ROA	−1.224 (−1.12)	−1.211 (−1.07)	0.066 (0.95)
Leverage	−1.803*** (−2.94)	−1.722*** (−2.72)	0.026 (0.47)
MB	0.183*** (4.19)	0.183*** (4.24)	0.003 (1.14)
RD	5.843*** (4.24)	5.296*** (3.65)	−0.126 (−1.08)
Stock return volatility	20.576** (2.13)	18.027* (1.83)	−2.556*** (−3.52)
Industry-adj. stock return	0.683** (2.37)	0.712** (2.45)	−0.012 (−0.51)
D(Insider)	−5.450*** (−7.02)	−5.391*** (−6.88)	0.005 (0.23)
CEO age	−0.033** (−2.22)	−0.027* (−1.74)	−0.004*** (−2.87)
Firm age	0.001	0.002	−0.001

(continued)

Table 8 (continued)

Dependent variable	Total pay		D(Severance)
	(1)	(2)	(3)
	(0.15)	(0.24)	(−1.19)
Constant	−5.609***	−7.426***	
	(−3.49)	(−4.30)	
Industry FE	YES	YES	YES
Year FE	YES	YES	YES
Observations	2,160	2,131	1,103
Adj. R-squared	0.383	0.386	
Adj. Pseudo R-squared			0.197

The symbol *** indicates that the p-value is less than 0.01, ** that it is less than 0.05, and * that it is less than 0.1

Overall, the regression results corroborate the prediction stated in Hypothesis 2. We find that job security has no value among CEOs who are risk-seeking (i.e. ambitious or turnaround experts in financially distressed firms). However, when CEOs are not risk-seeking, job security will have an impact on their compensation, and this impact will increase as CEOs become risk-adverse (i.e. external hires, outside-industry hires, or previous unemployed CEOs).

91.4.5 Job Security under Different Labor Market Conditions

We now investigate the link between job security and CEO pay under different labor market conditions.

Hypothesis 3 predicts that the price of job security increases with CEO's risk aversion; conversely, less risk averse CEOs demand lower job security related pay. To test this prediction, we separate our sample into two groups by industry labor market conditions. We categorize industry-years with above average demand/supply ratio as favorable labor market condition, and, vice versa, categorize industry-years with below average demand/supply ratio as poor labor market condition.

As shown in Table 9, job security under favorable labor market condition has no significant impact on CEO pay. This is because it is relatively easier for CEO candidate to find jobs under such market condition, and therefore the need of job security diminishes. On the other hand, job security has significant impact on CEO pay under poor market condition across all three job security measures. Overall, the regression results corroborate the prediction stated in Hypothesis 3. On average, under poor labor market condition, CEOs are compensated with \$39 thousand for each year decrease in average CEO tenure; \$479 thousand for CEOs with a predecessor who was not successfully retired; and 2.4% increase in the likelihood of demanding the easier incentive metric when the firm is unable to provide a greater pay.

Table 9 Regression Analysis for Firm-level Job Security under Different Labor Market Conditions. Column 1 to 4 report the OLS regression of CEO total pay across the three firm-level job security measures under different CEO labor market condition. We use Tobit regression for Time based grant/Total grant in Column 5 and 6 (left censored at 0 and right censored at 1). We separate our sample into two labor market condition groups based on the median of the CEO industry-year demand/supply ratio. The independent variables are defined in [Appendix A](#). Robust standard errors are reported in parentheses

Dependent variable	Total pay		Time based grant/Total grant			
	(1)	(2)	(3)	(4)	(5)	(6)
Market condition	Poor	Well	Poor	Well	Poor	Well
Firm_JS ₁	-0.039*** (-3.08)	0.026 (0.99)				
Firm_JS ₂			-0.479** (-2.13)	-0.422 (-1.27)		
Firm_JS ₃					0.024* (1.70)	0.026 (1.14)
Size	2.340*** (9.87)	2.160*** (12.48)	2.340*** (12.90)	2.165*** (12.51)	-0.052*** (-10.95)	-0.061*** (-9.12)
Sales growth	1.722*** (3.19)	-0.302 (-0.31)	1.881** (2.45)	-0.308 (-0.31)	0.002 (0.08)	0.093* (1.85)
ROA	-1.741 (-1.48)	0.040 (0.02)	-1.616 (-0.97)	-0.182 (-0.10)	-0.035 (-0.86)	-0.038 (-0.38)
Leverage	-1.995*** (-3.29)	0.219 (0.24)	-2.124** (-2.37)	0.100 (0.11)	-0.050 (-0.99)	-0.082 (-1.42)
MB	0.175** (2.30)	0.168** (2.25)	0.197*** (2.86)	0.165** (2.19)	-0.008*** (-3.73)	-0.003 (-1.34)
RD	5.563*** (4.58)	8.911*** (2.78)	6.341*** (4.34)	8.712*** (2.69)	-0.389*** (-6.54)	-0.525*** (-3.55)
Stock return volatility	34.811** (2.37)	-17.196 (-1.16)	37.408*** (3.01)	-19.534 (-1.34)	0.111 (0.17)	1.963** (2.06)
Industry-adj. stock return	0.818	0.427	0.779***	0.480	-0.040**	-0.060**

(continued)

Table 9 (continued)

Dependent variable	Total pay			Time based grant/Total grant		
	(1)	(2)	(3)	(4)	(5)	(6)
D(Insider)	(1.64) -2.856*** (-7.20)	(1.08) -2.803*** (-5.15)	(3.12) -2.828*** (-7.25)	(1.21) -2.733*** (-4.94)	(-2.02) 0.157*** (8.44)	(-2.42) 0.150*** (6.08)
CEO age	-0.033* (-2.00)	-0.052** (-2.47)	-0.030* (-1.86)	-0.055** (-2.58)	0.004*** (5.38)	0.005*** (3.29)
Firm age	0.002 (0.18)	-0.001 (-0.05)	0.001 (0.16)	0.000 (0.03)	0.001*** (3.55)	0.001 (1.49)
Constant	-10.036*** (-3.62)	-6.858*** (-3.60)	-9.378*** (-6.24)	-5.470*** (-2.87)	0.496*** (6.08)	0.551* (1.91)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	1,589	820	1,564	813	1,394	690
Adj. R-squared	0.353	0.408	0.353	0.406		
Pseudo R-squared					1.150	2.134

The symbol *** indicates that the p-value is less than 0.01, ** that it is less than 0.05, and * that it is less than 0.1

Table 10 Regression Analysis for Industry-level Job Security under Different Labor Market Conditions. Table 10 reports the OLS regression of CEO total pay across the three industry-level job security measures under different CEO labor market condition. We separate our sample into two labor market condition groups based on the median of the CEO demand/supply ratio. The independent variables are defined in [Appendix A](#). Robust standard errors are reported in parentheses

Dependent variable	Total Pay					
	(1)	(2)	(3)	(4)	(5)	(6)
Market condition	Poor	Well	Poor	Well	Poor	Well
Ind_JS ₁	-1.023** (-2.25)	-1.219 (-1.33)				
Ind_JS ₂			-0.584* (-1.84)	-1.450*** (-2.98)		
Ind_JS ₃					6.366*** (5.02)	3.603** (2.17)
Size	1.881*** (17.02)	2.348*** (12.87)	1.880*** (16.83)	2.313*** (12.87)	2.431*** (13.44)	2.693*** (9.57)
Sales growth	1.559* (1.80)	0.970 (1.05)	1.398 (1.60)	0.846 (0.93)	2.034 (1.42)	1.434 (0.99)
ROA	-0.941 (-0.71)	-0.557 (-0.39)	-1.033 (-0.77)	-0.759 (-0.54)	-3.349 (-0.91)	1.981 (0.62)
Leverage	-1.663** (-2.49)	-2.580*** (-2.30)	-1.647** (-2.43)	-2.147* (-1.89)	-2.606* (-1.96)	-1.705 (-1.12)
MB	0.207*** (4.75)	0.250** (2.47)	0.208*** (4.81)	0.235** (2.31)	0.187*** (2.79)	0.275** (2.06)
RD	9.409*** (6.12)	7.032** (2.53)	9.510*** (6.13)	7.231*** (2.66)	19.215*** (3.84)	5.615 (0.97)
Stock return	7.355 (0.67)	62.507*** (3.26)	8.530 (0.75)	56.665*** (3.01)	-24.489 (-1.29)	73.406** (2.21)
volatility						

(continued)

Table 10 (continued)

Dependent variable	Total Pay					
	(1)	(2)	(3)	(4)	(5)	(6)
Industry-adj.	-0.055	1.176**	-0.010	1.279***	0.247	1.868**
stock return	(-0.17)	(2.41)	(-0.03)	(2.63)	(0.43)	(2.33)
D(Insider)	-2.547***	-3.734***	-2.540***	-3.726***	-3.678***	-3.780***
	(-6.22)	(-6.25)	(-6.19)	(-6.21)	(-5.22)	(-4.38)
CEO age	-0.061***	-0.038	-0.061***	-0.033	-0.087**	-0.052
	(-3.27)	(-1.55)	(-3.31)	(-1.37)	(-2.52)	(-1.23)
Firm age	0.010	-0.021*	0.009	-0.016	0.002	-0.013
	(1.18)	(-1.72)	(1.13)	(-1.31)	(0.25)	(-0.95)
Constant	-3.816***	-8.615***	-3.651**	-8.595***	-13.409***	-17.008***
	(-2.67)	(-4.20)	(-2.56)	(-4.26)	(-4.18)	(-4.07)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	1,548	893	1,548	893	835	501
Adj. R-squared	0.321	0.332	0.319	0.335	0.279	0.310

The symbol *** indicates that the p-value is less than 0.01, ** that it is less than 0.05, and * that it is less than 0.1

We also examine the impacts of industry-level job security measure under different labor market condition categorized by the median of year average demand/supply ratio. We find that job security has a significant impact across all industry-level measures under poor labor market condition. Although some of the results also show that job security has a value even under well market condition, one of the measures, *Ind_JS₁*, becomes statistical insignificant. This is an evidence that job security's impact somewhat decreases under well market condition.

91.4.6 Robustness Test

We do not include corporate governance controls in our main analysis since it would dramatically decrease our sample size. However, the correlation between corporate governance and CEO compensation has been proven by many scholars (Core et al. 1999; Dicks 2012). In [Appendix B](#), we add corporate governance index (G-index), board independence (percentage of independent directors) and board size into our main models. At the same time, instead of measuring the total annual pay by the millions, we take the total pay reported on Execucomp (*tdc1*) by the natural log. Overall, the results remain in line with our main analysis.

We have proved that when firms are unable to grant the CEOs with greater pay, CEOs would seek for higher cash pay portion to compensate their job risks. However, this pay alternative might simply be the compensation scheme at financially distressed firms instead of firm's concession for job risk. For example, the high cash pay ratio can be merely caused by low valuation of firm stocks.

We give two explanations for this concern. First, low valuation of firm stock should not change CEO's cash pay ratio. CEOs are able to use prior firm performance to evaluate the firm and demand a favorable pay structure accordingly. When firm's stock price is low, CEO can, in return, demand more shares to compensate for such low valuation, and thus the cash pay ratio should remain the same. Second, if job security truly has a value in financially distressed firms, this value should increase as the job risk increases. We conduct a regression analysis by adding the interaction term of job security measure and financially distressed indicator (*Firm JS₃*). We construct a dummy variable equal to 1 when the predecessor of the CEO did not successfully retire, and thus indicating a risky job position. In [Appendix C](#), the results show that the interaction term is significantly positive, and therefore confirm that a more conservative pay structure in distressed firms is indeed caused by the decrease in job security.

91.5 Conclusion

We examine how job security has a price during CEO pay negotiation at hiring. We find significant impact of job security using different measures across firms. In general, CEO candidates and the board members would trade pay for job security; that is in terms of higher total pay. We further find this compensation for risk follow a

hierarchy order; when firms are unable to provide a higher pay, the CEOs will seek for alternative ways to compensate their job risk, including a severance agreement, a shorter grant vesting period, and an incentive plan metric that is easier to achieve. Moreover, job security has no value among risk-takers (i.e., ambitious CEOs or turnaround experts in financially distressed firms), and this price increases with CEO's risk aversion (i.e., external hires, outside industry hires, and previously unemployed CEOs). By examining the impact of job security under different market condition, we find that this premium disappears under favorable CEO labor market. Overall, our results are in line with the findings in labor economic literatures that workers are compensated in low job security positions. Our research compliments that of Peters and Wagner (2014), and is the first empirical study to examine the impact of CEO's risk preference and CEO market condition on the value of job security. We hope to add a new perspective to the determinants of CEO compensation and on the optimal design of managerial incentive contracts.

Appendix A: Variable Description

Variable Name	Description
Compensation	
Total pay (\$million)	Total annual compensation including salary, bonus and the grant date fair values of stock and option grants, long-term incentive payouts, and other compensation (tdc1 in Execucomp)
Total pay (scaled by size)	Total annual pay scaled by firm size
Cash/Total pay	Total cash components (salary + annual cash bonus) as a percentage of the total annual pay
Severance	Equals to 1 if the CEO signs a severance agreement
Avg. vesting period	The weighted average vesting period of grants
Time based/Total grant	Time based incentive grants over total grant
Job security	
Firm_JS ₁	Firm average CEO tenure before the new appointment
Firm_JS ₂	Equals to 1 if the outgoing CEO left the firm at a retiring age, and 0 otherwise
Firm_JS ₃	Equals to 1 for financially distressed firms, and 0 otherwise
Ind_JS ₁	The inverse coefficient of variation in industry stock returns
Ind_JS ₂	The inverse coefficient of variation in industry cash flow
Ind_JS ₃	The industry average credit rating
CEO Characteristics	
Ambitious	Equals to 1 if the CEO was from a rank lower than #3
D(Unemployeed)	Equals to 1 if the CEO was unemployed for 2 or more years
D(Insider)	Equals to 1 if the CEO was an internal hire
D(External)	Equals to 1 if the CEO was an external hire
D(Outside industry)	Equals to 1 if the CEO was an outside industry hire
D(Expert)	Equals to 1 if the CEO is a workout expert

(continued)

Variable Name	Description
D(Unsuccessful outgoing CEO)	Equals to 1 if the predecessor was not successfully retired
Market condition	
Demand/supply	The total CEO job demand divide by the total CEO supplies in a given year
Industry demand/supply	The total CEO job demand divide by the total CEO supplies within the industry in a given year
Controls	
lnSize	Natural logarithm of total asset
Sales growth	Logarithm of the ratio between sales and lagged sales
ROA	Operating income divided by total assets
Leverage	The sum of long-term debt and debt in current liabilities divided by total assets
M/B	Market value of total assets over book value of total assets
R&D	Research and development expense divided by total assets
Stock volatility	The standard deviation of daily stock returns in each year
Industry-adj. stock return	Annual stock return minus its industry mean.
G-index	Gompers, Ishii, and Metrick (2003) governance Index
Board independence	Percentage of independent directors on board
Board size	Number of board members

Appendix B: Robustness Tests for the General Case

Column 1 to 5 report the results of OLS regression. The dependent variable is the neutral log of the total pay reported in the Execucomp database. Robust standard errors are reported in parentheses. We add corporate governance index (G-index), board independence (percentage of independent directors) and board size into our main models as control variables.

Dependent Variable:	Industry-level			Firm-level	
Total pay	(1)	(2)	(3)	(4)	(5)
Ind_JS ₁	−0.114** (−2.98)				
Ind_JS ₂		−0.184*** (−3.55)			
Ind_JS ₃			1.031*** (6.75)		
Firm_JS ₁				−0.002 (−0.68)	
Firm_JS ₂					−0.090** (−2.45)
Size	0.394***	0.397***	0.400***	0.445***	0.447***

(continued)

Dependent Variable:	Industry-level			Firm-level	
Total pay	(1)	(2)	(3)	(4)	(5)
	(9.50)	(23.34)	(16.54)	(21.44)	(20.10)
Sales growth	0.066	0.050	0.024	0.042	0.080
	(0.43)	(0.45)	(0.17)	(0.30)	(0.73)
ROA	0.614	0.800***	0.509	0.645	0.705**
	(1.23)	(3.08)	(1.27)	(1.75)	(2.26)
Leverage	0.080	0.131	−0.308*	0.020	−0.001
	(0.32)	(0.95)	(−1.67)	(0.11)	(−0.01)
MB	0.038**	0.038***	0.034***	0.032***	0.033***
	(2.91)	(5.21)	(3.46)	(3.40)	(3.66)
RD	2.327***	2.250***	2.888***	1.816***	1.739***
	(7.84)	(7.04)	(5.30)	(6.34)	(6.03)
Stock return volatility		5.780***	2.063	5.745**	5.415
		(2.75)	(0.77)	(2.27)	(1.73)
Industry-adj. stock return	0.168*	0.187***	0.161**	0.213***	0.222***
	(2.16)	(3.43)	(2.11)	(3.18)	(9.93)
D(insider)	−0.524***	−0.515***	−0.512***	−0.497***	−0.484***
	(−6.25)	(−8.31)	(−6.41)	(−6.97)	(−8.66)
CEO age	−0.017***	−0.017***	−0.018***	−0.016***	−0.016***
	(−11.29)	(−4.71)	(−3.51)	(−8.64)	(−7.60)
Firm age	−0.001	−0.000	−0.001	−0.002**	−0.002
	(−0.72)	(−0.19)	(−0.62)	(−2.44)	(−1.54)
G-index	0.017	0.016	0.017	0.020**	0.022*
	(1.93)	(1.28)	(1.07)	(2.20)	(2.09)
Board independence	−0.694	−0.597	−0.114	−0.417	−0.453
	(−1.16)	(−1.45)	(−0.19)	(−0.83)	(−0.96)
Board size	−0.018	−0.015	−0.011	−0.005	−0.007
	(−0.96)	(−1.38)	(−0.82)	(−0.46)	(−0.56)
Constant	6.303***	5.977***	4.808***	5.696***	5.683***
	(50.52)	(22.73)	(12.43)	(30.34)	(22.84)
Observations	1,638	1,638	972	1,638	1,617
Industry FE	NO	NO	NO	YES	YES
Year FE	YES	YES	YES	YES	YES
Adj. R-squared	0.415	0.422	0.365	0.462	0.465

The symbol *** indicates that the p-value is less than 0.01, ** that it is less than 0.05, and * that it is less than 0.1

Appendix C: Robustness Tests for Financially Distressed Firms

The following table report the results of Tobit regression (left censored at 0 and right censored at 1). The dependent variable is cash pay ratio. *D(Unsuccessful outgoing CEO)* is an indicator that equal to one if the outgoing CEO was not successfully

retired. In Column 2, we add corporate governance index (G-index), board independence (percentage of independent directors) and board size into our main models as control variables.

Dependent variable:	(1)	(2)
Cash/Total pay		
Firm_JS ₃ *D(unsuccessful outgoing CEO)	0.039*** (4.46)	0.046** (2.48)
Firm_JS ₃	0.007 (0.94)	0.020 (1.28)
D(unsuccessful outgoing CEO)	-0.022*** (-3.60)	-0.019** (-2.07)
Size	-0.056*** (-14.83)	-0.058*** (-14.36)
Sales growth	0.034 (1.17)	0.062 (1.32)
ROA	-0.050 (-0.89)	-0.043 (-0.66)
Leverage	-0.050*** (-3.22)	-0.052 (-1.10)
MB	-0.007*** (-5.49)	-0.007*** (-3.79)
RD	-0.431*** (-15.52)	-0.381*** (-3.43)
Stock return volatility	0.409 (0.56)	-0.123 (-0.13)
Industry-adj. stock return	-0.046*** (-3.48)	-0.054*** (-4.55)
D(insider)	0.152*** (12.34)	0.144*** (7.78)
CEO age	0.004*** (4.62)	0.005*** (5.18)
Firm age	0.001*** (5.42)	0.001*** (5.30)
G-index		-0.006 (-1.40)
Board independence		0.256* (1.71)
Board size		-0.000 (-0.05)
Constant	0.549*** (7.61)	0.609*** (7.76)
Observations	2,089	1,412
Industry FE	YES	YES

(continued)

Dependent variable:	(1)	(2)
Year FE	YES	YES
Adj. R-squared	1.131	1.856

The symbol *** indicates that the p-value is less than 0.01, ** that it is less than 0.05, and * that it is less than 0.1

References

- Abdel-Khalik, A. Rashad. 2003. Self-sorting, incentive compensation and human-capital assets. *The European Accounting Review* 12: 661–697.
- Almazan, Andres, and Javier Suarez. 2003. Entrenchment and severance pay in optimal governance structures. *Journal of Finance* 58: 519–547.
- Altman, Edward I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23: 589–609.
- Baber, William R., Surya N. Janakiraman, and Sok-Hyon Kang. 1996. Investment opportunities and the structure of executive compensation. *Journal of Accounting and Economics* 21: 297–318.
- Bacharach, Samuel B., and Edward J. Lawler. 1981. *Bargaining: Power, tactics and outcomes (ERIC)*. San Francisco: Jossey-Bass.
- Bena, Jan, Hernan Ortiz-Molina, and Elena Simintzi. 2020. Shielding firm value: employment protection and process innovation. Available at SSRN 3223176.
- Bertola, Giuseppe. 1990. Job security, employment and wages. *European Economic Review* 34: 851–879.
- Böckerman, Petri, Pekka Ilmakunnas, and Edvard Johansson. 2011. Job security and employee well-being: Evidence from matched survey and register data. *Labour Economics* 18: 547–554.
- Cadman, Brian, Sandy Klasa, and Steve Matsunaga. 2010. Determinants of ceo pay: A comparison of execucomp and non-execucomp firms. *The Accounting Review* 85: 1511–1543.
- Cadman, Brian D., Tjomme O. Rusticus, and Jayanthi Sunder. 2013. Stock option grant vesting terms: Economic and financial reporting determinants. *Review of Accounting Studies* 18: 1159–1190.
- Campbell, David, Alan Carruth, Andrew Dickerson, and Francis Green. 2007. Job insecurity and wages*. *The Economic Journal* 117: 544–566.
- Coles, Jeffrey L., Naveen D. Daniel, and Lalitha Naveen. 2006. Managerial incentives and risk-taking. *Journal of Financial Economics* 79: 431–468.
- Core, John E., Robert W. Holthausen, and David F. Larcker. 1999. Corporate governance, chief executive officer compensation, and firm performance¹. *Journal of Financial Economics* 51: 371–406.
- Dai, Yunhao, P. Raghavendra Rau, Aris Stouraitis, and Weiqiang Tan. 2020. An ill wind? Terrorist attacks and CEO compensation. *Journal of Financial Economics* 135: 379–398.
- DeFond, Mark L., and Chul W. Park. 1997. Smoothing income in anticipation of future earnings. *Journal of Accounting and Economics* 23: 115–139.
- Demsetz, Harold, and Lehn, Kenneth. 1985. The structure of corporate ownership: Causes and consequences. *Journal of political economy* 93.6: 1155–1177.
- Desai, Hemang, Chris E. Hogan, and Michael S. Wilkins. 2006. The reputational penalty for aggressive accounting: Earnings restatements and management turnover. *The Accounting Review* 81: 83–112.
- Dicks, David L. 2012. Executive compensation and the role for corporate governance regulation. *Review of Financial Studies* 25: 1971–2004.
- Ellis, Jesse A. 2012. Are turnaround specialists special? An examination of ceo reputation and ceo succession. Available at SSRN 1956691.

- Engel, Ellen, Rachel M. Hayes, and Wang Xue. 2003. Ceo turnover and properties of accounting information. *Journal of Accounting and Economics* 36: 197–226.
- Finkelstein, Sydney, and D. Hambrick. 1997. Review: Strategic leadership: Top executives and their effects on organizations. *Australian Journal of Management* 22: 221–224.
- Fudenberg, Drew, and Jean Tirole. 1995. A theory of income and dividend smoothing based on incumbency rents. *Journal of Political Economy* 103: 75–93.
- Gibbons, Robert, and Kevin J. Murphy. 1992. Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy* 100: 468–505.
- Gilson, Stuart C. 1989. Management turnover and financial distress. *Journal of Financial Economics* 25: 241–262.
- Gopalan, Radhakrishnan, Todd Milbourn, Fenghua Song, and Anjan V. Thakor. 2014. Duration of executive compensation. *The Journal of Finance* 69: 2777–2817.
- Göx, Robert F., and Thomas Hemmer. 2020. On the relation between managerial power and CEO pay. *Journal of Accounting and Economics* 69: 101300.
- Harris, Dawn, and Constance Helfat. 1997. Specificity of ceo human capital and compensation. *Strategic Management Journal* 18: 895–920.
- Hermalin, Benjamin E., and Michael S. Weisbach. 2001. Boards of directors as an endogenously determined institution: A survey of the economic literature. *Journal of Financial Economics* 96: 195–214.
- Jenter, Dirk and Kanaan, Fadi. 2021. CEO Turnover and Relative Performance Evaluation. Available at SSRN 888270
- Kaplan, Steven N., and Bernadette A. Minton. 2012. How has ceo turnover changed? *International Review of Finance* 12: 57–87.
- Kelley, Harold H., Linda Linden Beckman, and Claude S. Fischer. 1967. Negotiating the division of a reward under incomplete information. *Journal of Experimental Social Psychology* 3: 361–398.
- Kempf, Alexander, Stefan Ruenzi, and Tanja Thiele. 2009. Employment risk, compensation incentives, and managerial risk taking: Evidence from the mutual fund industry. *Journal of Financial Economics* 92: 92–108.
- Kini, Omesh, Ryan Williams, and David Yin. 2020. CEO non-compete agreements, job risk, and compensation. *The Review of Financial Studies* 34: 4701–4744.
- Klein, Felice B., Pierre Chaigneau, and Cynthia E. Devers. 2021. CEO gender-based termination concerns: Evidence from initial severance agreements. *Journal of Management* 47: 567–596.
- Lazear, Edward P. 1990. Job security provisions and employment. *The Quarterly Journal of Economics* 105: 699–726.
- . 1998. *Personnel economics for managers*. New York: Wiley.
- Low, Angie. 2009. Managerial risk-taking behavior and equity-based compensation. *Journal of Financial Economics* 92: 470–490.
- Mannix, Elizabeth A., Leigh L. Thompson, and Max H. Bazerman. 1989. Negotiation in small groups. *Journal of Applied Psychology* 74: 508–517.
- Manso, Gustavo. 2011. Motivating innovation. *The Journal of Finance* 66: 1823–1860.
- Meindl, James R., Sanford B. Ehrlich, and Janet M. Dukerich. 1985. The romance of leadership. *Administrative Science Quarterly* 30: 78–102.
- Nguyen, Pascal. 2007. Macroeconomic factors and Japan's industry risk. *Journal of Multinational Financial Management* 17: 173–185.
- Olaniyi, C. 2019. Asymmetric information phenomenon in the link between CEO pay and firm performance: An innovative approach. *Journal of Economic Studies* 46: 306–323.
- Oyer, Paul. 2004. Why do firms use incentives that have no incentive effects? *The Journal of Finance* 59: 1619–1650.
- Parrino, Robert. 1997. Ceo turnover and outside succession a cross-sectional analysis. *Journal of Financial Economics* 46: 165–197.
- Peters, Florian S., and Alexander F. Wagner. 2014. The executive turnover risk premium. *The Journal of Finance* 69: 1529–1563.

- Pfeffer, Jeffrey, and Cathy Lammerding. 1981. *Power in organizations*. Boston, MA: Pitman Marshfield.
- Pinkley, Robin L., Margaret A. Neale, and Rebecca J. Bennett. 1994. The impact of alternatives to settlement in dyadic negotiation. *Organizational Behavior and Human Decision Processes* 57: 97–116.
- Rau, P. Raghavendra, and Jin Xu. 2013. How do ex ante severance pay contracts fit into optimal executive incentive schemes? *Journal of Accounting Research* 51: 631–671.
- Sebora, Terrence C. 1996. Ceo-board relationship evaluation: An exploratory investigation of the influence of base rate factors. *Journal of Managerial Issues* 8: 54–77.
- Wade, James B., Joseph F. Porac, Timothy G. Pollock, and Scott D. Graffin. 2006. The burden of celebrity: The impact of ceo certification contests on ceo pay and performance. *Academy of Management Journal* 49: 643–660.
- Ward, Andrew, Jeffrey A. Sonnenfeld, and John R. Kimberly. 1995. In search of a kingdom: Determinants of subsequent career outcomes for chief executives who are fired. *Human Resource Management* 34: 117–139.
- Zemba, Yuriko, Maia J. Young, and Michael W. Morris. 2006. Blaming leaders for organizational accidents: Proxy logic in collective- versus individual-agency cultures. *Organizational Behavior and Human Decision Processes* 101: 36–51.



Tail-Risk Protection: Machine Learning Meets Modern Econometrics

92

Bruno Spilak and Wolfgang Karl Härdle

Abstract

Tail risk protection is in the focus of the financial industry and requires solid mathematical and statistical tools, especially when a trading strategy is derived. Recent hype driven by machine learning (ML) mechanisms has raised the necessity to display and understand the functionality of ML tools. In this chapter, we present a dynamic tail risk protection strategy that targets a maximum predefined level of risk measured by Value-At-Risk while controlling for participation in bull market regimes. We propose different weak classifiers, parametric and nonparametric, that estimate the exceedance probability of the risk level from which we derive trading signals in order to hedge tail events. We then compare the different approaches both with statistical and trading strategy performance, finally we propose an ensemble classifier that produces a meta tail risk protection strategy improving both generalization and trading performance.

Keywords

Tail-risk · Trading strategy · Cryptocurrency · Value-At-Risk · Deep learning · Machine learning · Econometrics · Extreme value theory · Exceedance probability

B. Spilak

IRTG 1792, School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany
e-mail: bruno.spilak@hu-berlin.de

W. K. Härdle (✉)

Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

Singapore Management University, Singapore Department of Mathematics and Physics, Singapore, Singapore

Charles University in Prague, Prague, Czech Republic

National Yang Ming Chiao Tung University (NYCU), Hsinchu City, Taiwan
e-mail: haerdle@hu-berlin.de

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_94

2177

Abbreviations	
DL	Deep Learning
ML	Machine Learning
VaR	Value-at-Risk
EVT	Extreme Value Theory
LPA	Local Parametric Approach
MLP	Multilayer Perceptron
LSTM	Long Short-Term Memory
GPD	Generalised Pareto distribution
QMLE	Quasi-Maximum Likelihood Estimation
AUC	Area Under the roc Curve

92.1 Introduction

“Black swan events” are occurring from time to time. After 2008, investors in the DAX had to wait 5 years before recovering their loss. The Nikkei stock index still has not yet recovered, even 25 years after the bubble burst. The cryptocurrency market has plunged drastically at the end of 2017. What choices are left to investors in these situations?

Indeed, it is one of the main challenges of quantitative finance to build models and develop tools able to protect investments from such extreme events (Bollerslev and Todorov 2011). The first risk management strategy addressing the latter is diversification but Longin and Solnik (2001) show that when financial markets exhibit huge downturn periods, correlation significantly increases, hence are counteracting the effect of portfolio diversification. Another strategy for tail risk protection is option based, where the risk manager is long in put options. Yet, loss aversion leads to high prices of put options (Kozhan et al. 2013; Packham et al. 2017). A solution to escape this dilemma is to build a strategy with dynamic asset allocation. The risk manager creates an asymmetric risk profile with participation in upside market and protection against severe loss. This trading strategy is often called tail risk protection strategy by practitioners and is much more cost efficient and flexible than the options strategy (Franke et al. 2019). In order to pursue such a strategy, one needs to predict distributional properties of the portfolio. This, of course, is obviously impossible since not only distributions but also their parameters change in time. This motivates our entry where we hint on comparing Econometrics tools, Machine learning (ML), and modern Local Parametric Approach (LPA). We present a dynamic tail risk protection strategy that targets a maximum predefined level of risk measured by Value-At-Risk while controlling for participation in bull market regimes. We propose different weak classifiers, parametric, based on GARCH tool, and nonparametric using ML, that estimate the exceedance probability of the risk level from which we derive trading signals in order to hedge tail events.

The GARCH tool with normally distributed innovations allows us to catch volatility clusters and yields good volatility forecasts. The application of Extreme

Value Theory (EVT) to GARCH residuals provides insight into the tail probabilities, that is, the likeliness of an extreme loss. Nevertheless, econometric models rely on strong assumptions and cannot deal with structural breaks that happen in financial time series.

ML has been successful in many applications during the last years thanks to its generalization power on large datasets, but, in quantitative finance, in particular for financial returns forecasting, ML did not prove its superiority in comparison with its application to text analysis or image recognition, where Deep Learning (DL) tools became state-of-the-art. Moreover, the black box effect of DL makes the industry reluctant to invest in such models. In theory, DL can extract information from nonlinear relations in high dimensional space, so for quantitative finance practitioners, it should be natural to use the machine learning tool box. Nevertheless, DL models suffer from overfitting and can be very difficult to train. If they offer similar prediction accuracy than econometrics, it might be preferable to use the latter with regards to its interpretability, lower complexity, and cheaper computational costs.

The cryptocurrency market has experienced an exponential growth during 2017, where BTC peaked at \$19783.06 on December 17, 2017, and dropped below \$14,000 24 h later, losing one third of its value. One year later, on December 7, 2018, BTC price briefly dipped below \$3300, a 76% drop from the previous year and a 15-month low. At the time of writing, the cryptocurrency market is experiencing a new large uptrend as swathes of institutional investors are gaining new interest because of the launch of new futures contracts on both regulated and unregulated cryptocurrency exchanges. Central banks also reacted to private initiatives of launching digital currencies, such as Facebook's Libra, as they may have the potential to dilute the main power of central banks – to control the supply of money to the economy. With high volatility and new interest in the cryptocurrency market, the BTC ecosystem is a perfect environment to show the effectiveness of tail risk strategy. Indeed, its high volatility promises high gains in upward movement of the price. Our goal is then to protect investors from large downturns of the market through an accurate prediction of tail risk.

Since Basel committee on Banking Supervision Amendment to incorporate market risk (1996), regulators imposed the use of certain metrics to measure the risk of investments, such as Value- At-Risk (VaR) and Expected-Shortfall (ES). In order to meet both investors' and regulators' will, we build a tail risk protection strategy that controls the VaR of our portfolio by ensuring that it is below a certain level, denoted as target VaR.

As high volatility segments often precede market swings, the GARCH model is a natural tool for tail risk protection, since GARCH catches these volatility clusters. Nevertheless, a volatility-based risk management strategy forces us to divest in such period, reducing alpha possibilities in case of positive movement. Since financial returns have heavy tails, the EVTGARCH McNeil and Frey (2000) allow us to improve our forecast of the tail event direction. Indeed, seeking alpha, excess return over a benchmark, has become more challenging for banks and institutional investors which follow strong regulations based on VaR estimates. We address this problem here by focusing on dynamic risk management based on econometrics,

explaining stylized facts of financial time series with parametric approaches, and non- or local parametric methods, in particular the LPA from Spokoiny (2009), Multilayer Perceptron (MLP), and Long Short-Term Memory (LSTM) neural networks, which catches nonlinear features with memory.

Our contribution is a rigorous comparison of GARCH- and ML-based tools in the context of extreme loss prediction based on their forecast in a classification scheme. On top, we show how to use standard Ensemble method as a meta classifier that produces a hybrid tail risk protection strategy, improving both generalization and trading performance by taking advantage of each approach. Finally, we evaluate our strategy with a realistic backtest including trading fees, by comparing it with classical buy-and-hold benchmark and other recent machine learning oriented tail risk protection strategies, such as the constant target VaR from Rickenberg (2019) and the Varspread strategy from Packham et al. (2017).

The results showed here will certainly motivate practitioners to apply ML techniques in order to improve GARCH performance. We provide comparison metrics such as forecast error, classification metrics, and backtest results of our tail risk strategies on the cryptocurrency market with BTC investment. We also provide robustness checks through cross-validation.

This chapter is structured as follows. First we review the current literature of our subject, then we explain the trading strategy we aim to build. In the third and fourth part, we explain the theoretical models used to build our strategy. In the final section, we present our results.

92.2 Background and Literature Review

92.2.1 Volatility as a Risk Measure?

A good risk measure must be tailored to the investor's preferences which are often unknown in practice. Starting from preferences, practitioners often make assumptions in order to build an "optimal risk measure," corresponding to an imagined investor's goal. A large part of the financial literature studies volatility as a risk measure, since it is nicely tied to Gaussian and LS techniques. For example, the VIX, referred to as the "investor fear gauge," is often taken as a sentiment indicator since volatility reflects investors' aversion to risk. The mean-variance portfolio, developed by Markowitz, is built under such assumptions where the weights of the risky assets included in the portfolio are derived from their volatility. Nevertheless, financial returns often have fat-tails and are not normally distributed, which Markowitz theory does not account for. Finally, volatility, being symmetric, is not realistic as a risk measure, since it does not take into consideration investors' loss aversion, weighting equally volatility associated with gains and losses. Most investors are more concerned about downside risk, or losses, rather than volatility (Bollerslev et al. 2015). In this chapter, our goal is to build a trading strategy avoiding large losses or tail risk which is better suited for loss averse investors.

Packham et al. (2017) showed that a trading strategy accounting for tail risk can outperform simple buy-and-hold and traditional portfolio protection strategies. By using models such as GARCH with normally distributed innovations and GARCH with innovations following a Generalized Pareto Distribution (GPD), they built a new criterion for riskiness defined as the evolution of the estimated Value-At-Risk (VaR) spread between the two models. Thus, in period of increasing tail risk, this spread is significantly different from 0 which allows the trader to take adequate decisions. We denote this strategy as the Varspread strategy. Rickenberg (2019) compares different risk measure such as volatility, VaR, and Conditional-Value-at-Risk (CVaR), also named Expected Shortfall (ES), in order to build dynamic trading strategies and find that downside risk measures outperform volatility in terms of a higher Sharpe Ratio, better drawdown protection, and higher utility gains for mean-variance and loss-averse investors. Happersberger et al. (2019) also focus on ES and VaR forecasts in order to manage dynamic tail risk protection strategies.

All the papers mentioned above and in general the literature of tail risk protection focus on predicting risk measures in a regression manner and study the total distribution of the returns where the goal is to minimize Mean Squared Error (MSE) type measures of fit. Our argument is that it is not necessary to predict the total distribution of the return, whether it is crucial to correctly forecast the direction of the tail, in particular since in our case big profits or losses are at stake (Jordà and Taylor 2011).

92.2.2 Investors' Preferences

Rickenberg (2019) developed a tail risk protection strategy, denoted target VaR strategy, where the trading signals are calibrated so the VaR of the strategy is constant over time for a predefined significance level α . The weights of the risky asset in the simple portfolio consisting of two assets, one risky and one risky-less asset, is a function of a constant level of risk defined as a VaR level denoted, target VaR, depending on investors' risk aversion measured by α . The weights are calibrated so the VaR of portfolio is constant and equals to the target VaR. The target VaR strategy has the main advantage to be better interpretable for investors who can prescribe their acceptable loss limit to the trader. Nevertheless, it is assumed that investors' preferences are static since the target VaR is fixed and does not depend on t . Such strategy is based on the standard financial theory assumption that investors are rational and have invariant risk preferences.

However, with the development of Behavioral Finance, numerous studies draw attention on the fact that investors are often irrational and their preferences change with different situations. The early paper of Kahneman and Tversky (1979) shows that investors are more risk averse with gains, but less with losses. More recently, Wen et al. (2014) study the characteristics of investors' risk preferences with different states of gains and losses and show that investors' risk preferences are time-varying with them. Indeed, the degree of risk aversion rises with the increasing gains and that of risk seeking improves with the increasing losses.

Taking into consideration investors' time-varying preferences, a constant target VaR strategy is inadequate. Indeed, we should decrease our target VaR in period of gains and increase it in periods of losses to allow more conservative trades in period of gains and more aggressive ones in case of losses.

92.2.3 Dynamic Tail Risk Protection Strategy

Our goal is to build a dynamic tail risk protection with an adaptive predefined level or risk, adapting to the benchmark's performance which influences investors' preferences (Wen et al. 2014). At constant significance level α , when the benchmark is experiencing a period of good performance where the gains are increasing, its VaR is decreasing; thus, the investors give more attention to smaller losses and tend to be more risk averse. As follows, the trader should adapt its strategy and aim for a smaller target VaR. In the inverse situation, during a period of losses, the tail risk of the benchmark is increasing and the investor only pays attention to large losses, tending to seek more risk. For example, one could use a threshold as a function of volatility in order to build a dynamic threshold labeling function as in de Prado (2018), chapter 3.3. In the next section, we will explain how to build such strategy.

92.3 Trading Strategy

92.3.1 Tail Risk as a Risk Measure

Let us give the following **tail risk definition**. Throughout this chapter, we consider a risky asset, for example, one stock, with price process p_t , where $t \in [0, \dots, T]$ with T the final period time step. As usual, we define the one period log-return, where one period is here 1 h, as $r_t = \log \frac{p_t}{p_{t-1}}$ and the loss series $-r_t$. As in Packham et al. (2017), we define the Value-At-Risk for a risk level α , denoted VaR_t^α , as the α -quantile of the distribution of the loss $-r_t$:

$$\text{VaR}_t^\alpha = \inf \{l \in \mathbb{R} : \mathbb{P}(-r_t > l) \leq 1 - \alpha\} = \inf \{l \in \mathbb{R} : \mathbb{P}(-r_t \leq l) \geq \alpha\}$$

If r_t follows an absolutely continuous loss distribution, then the Value-At-Risk can be defined as an unlikely and severe loss which satisfies:

$$\mathbb{P}(-r_t \leq \text{VaR}_t^\alpha) = 1 - \alpha$$

where α is small corresponding to investors' risk aversion. In practice, the risk level often takes the value of 0.01, 0.025, or 0.05. The Value-At-Risk characterizes the far right tail of the distribution of the loss; thus, we use it as a measure of tail risk. For example, a $\text{VaR} = u_t^\alpha = 0.05$ for $\alpha = 0.01$ means that returns below -5% only happen 1% of the time. Moreover, for any threshold u , $\mathbb{P}(-r_t \geq u)$ is called the exceedance probability over u .

92.3.2 Tail Risk Protection Strategy

Now, let us consider a simple buy/sell trading strategy where the trader decides at time t either to enter the market with a full position (he invests the totality of his available capital at time t into BTC) or to stay out of the market (he sells all available BTC or does not invest capital at time t in order to have all his capital in a risk free asset) based on the information available at time t denoted $\mathcal{F}_t$. The trader's decision can be represented as a binary variable, or trading signal, $s_t \in \{0, 1\}$ where 1 corresponds to the decision to stay out of the market and 0 to the decision to fully invest the capital at time t . The return of such strategy is defined for all $t \in [1, \dots, T]$ as $R_t = (1 - s_t - 1)r_t$ and the excess return $R_t - r_t$.

Target VaR The goal of tail risk protection is to maximize the expected economic utility of a risk averse investor which can be represented by the risk-adjusted return of the strategy, where risk is characterized by the tail behavior of the portfolio. In other words, our goal is to lower the probability of tail risk, that is, at constant level T , to have a lower portfolio VaR than the buy-and-hold benchmark strategy which buys the asset at the beginning of the period and sell it at time T , the end of the period of investment. In parallel, we have two choices to maximize the return, either we can maximize the expected return, following the portfolio selection criterion from Markowitz (1952), or we can maximize the total return at T , as suggested by Kelly (1956). Since in our setting, we are optimizing the strategy for multiple periods (Hakansson 1971) we choose the latter. We can write the following optimization program:

$$\begin{aligned} \max \quad & \sum_{t=1}^T (R_t - r_t) \\ \text{s.t.} \quad & \text{VaR}_t^\alpha \leq \text{TVaR}_t^\alpha \end{aligned} \quad (1)$$

where VaR^α is the portfolio VaR and TVaR_t^α is the **target VaR for level α** given by the investor.

We have the following equivalence, $\forall t \in [1, \dots, T]$:

$$\begin{aligned} \text{VaR}_t^\alpha &\leq \text{TVaR}_t^\alpha \\ \text{P}(-R_t \leq \text{TVaR}_t^\alpha) &\geq 1 - \alpha \end{aligned} \quad (2)$$

Thus, one relaxes the constant constraint from Rickenberg (2019) and aims at constructing a trading mechanism with signal $s = \{s_1, \dots, s_T\}$ that has a maximum TVaR^α . From now on, we think of fixed α and we will write TVaR_t^α as TVaR_t for the benchmark Value-At-Risk and VaR_t^α as VaR_t for the portfolio.

How to define the signals s_t For all $t \in [0, T-1]$, we define:

$$s_t = \mathbf{I}_{-r_{t+1} \geq \text{TVaR}_{t+1}} \quad (3)$$

and by construction $R_t = (1 - \mathbf{I}_{-r_t \geq \text{TVaR}_t})r_t$. Since TVaR_t is strictly positive for financial assets loss series, we have $\text{P}(-R_t \leq \text{TVaR}_t) = 1$, thus (2) is verified.

Since we do not know the true distribution of r_t and, obviously, at time t , we do not observe r_{t+1} , we do not know s_t , and we must build estimates based on the observation available in order to make a trading decision. We estimate the following conditional probability, $P(-r_{t+1} \geq \text{TVaR}_{t+1} | \mathcal{F}_t)$, which is the exceedance probability over the threshold TVaR_{t+1} , and decide whether the trader must close his position $s_t = 1$ or stay in the market $s_t = 0$.

Tail Risk Target TVaR^α The tail risk target, TVaR^α , is defined by the investors' preferences. Here, our goal is not to study investors' preferences, but to show how to build trading signals based on estimates of the exceedance probability over a given risk level. This problem has been well studied in the literature, for various applications, such as seismic risk assessment (Honegger and Wijewickreme 2013), risk assessment for decision making with application to terrorism (Kunreuther 2002), or again floods, earthquakes, drought, and hurricanes risk management (Lambert et al. 1994; Mason et al. 2007). For trading strategies, Christoffersen and Diebold (2006) and Linton and Whang (2007) considered exceedances above 0, whereas Chung and Hong (2007) focused on nonzero thresholds. In particular, Taylor and Yu (2016) studied thresholds that are not close to 0, as this is of greater relevance for risk management. This is also our interest but we consider extreme thresholds which are time varying and can be used as tail risk measure.

For illustration, we focus in this chapter on different tail risk targets derived from the data itself. That is we use sample quantiles for different level α computed on the historical losses in a rolling window manner for different window size. In other words, to describe the investors' preferences for the next trading period $t + 1$, we use the historical VaR of the loss series as tail risk measure defined for $w \in [1, T]$ as:

$$\forall t \in [w, T], \forall i \in [t - w + 1, t], \text{TVaR}_{t+1} = \text{hist-VaR}_t^{\alpha, w} = \inf \{l \in \mathbb{R} : P(-r_i > l) \leq 1 - \alpha\} \quad (4)$$

The main advantage of this risk target is its computational simplicity allowing us to easily build training labels for our classifiers that we will develop in the next sections. On top, it is adaptable to investors' preferences, since using a small rolling window would suit an investor with varying preferences where the risk target quickly adapts itself to changes in the true return distribution, while large rolling windows give more stable risk target, corresponding to an investor with static preferences. In this chapter, we will use three window sizes: 24 (1 day), 2880 (4 months), and 4320 (6 months), since we are using hourly data, which are plotted on Fig. 1 for $\alpha = 0.01$. We give more details about the dataset in the next section. Our choice is motivated by the goal to study how the tools used here react to different levels of noise in the target variable.

As we can see in Fig. 1, the $\text{TVaR}_t^{0.01, 24}$ is much more conservative than the $\text{TVaR}_t^{0.01, w}$ with larger w , with much more losses exceeding the risk target. Indeed, we have, respectively, 0.045, 0.012, and 0.011 exceedance for the $\text{TVaR}_t^{0.01, w}$ with w is 24, 2880, and 4320, where the exceedance is defined as $N_e = \frac{1}{T} \sum_{t=1}^T -r_t \geq \text{TVaR}_t^\alpha$, that is the proportion of data exceeding the threshold. A strategy with

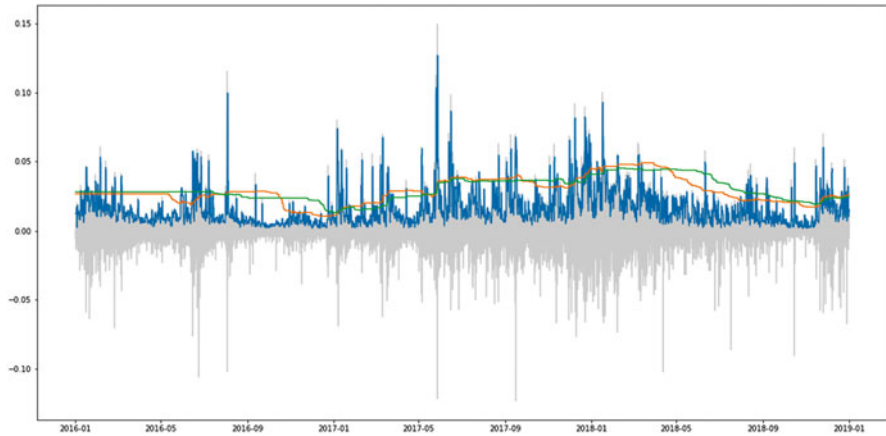


Fig. 1 Hourly btc losses and $TVaR_t^{0.01,w}$ for different window size, $w = 24$ (1 day), $w = 2880$ (4 months), $w = 4320$ (6 months)

$TVaR_t^{0.01,24}$ as tail risk target should be much more conservative and hedges much more losses than the two other risk targets.

Trader's Decision Finally, how to decide whether we should hedge our position for the next trading period? We could build an estimation of the return at the next period, r_{t+1} , in a regression manner, nevertheless, we are only interested in the distribution of the tail. Thus, in this chapter, we directly approximate the conditional probability classes of the variable s_t , that is p_t , defined $\forall t \in [0, \dots, T-1]$ as:

$$p_t = P(s_t = 1 | \mathcal{F}_t) = P(-r_{t+1} \geq TVaR_{t+1}^\alpha | \mathcal{F}_t) \quad (5)$$

From p_t , we can make the trading decision $\hat{s}_t$ defined for a threshold $u \in \mathbb{R}$:

$$\hat{s}_t = \begin{cases} 1, & \text{if } p_t \geq u \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

We can now write the realized return of our strategy, for all $t \in [1, T]$:

$$R_t = (1 - \hat{s}_{t-1})r_t \quad (7)$$

92.4 Machine Learning Trader

Let $X_t \in \mathbb{R}^p$ be some input feature vector of dimension p which summarize $\mathcal{F}_t$. The aim is to approximate the true unknown conditional probability $P(s_t | X_t)$ with a learning algorithm $\mathcal{M}$ for different risk targets. We consider four different α : 1%,

2.5%, 5%, and 10%, and three rolling windows to compute TVAR: 24, 2880, and 4320. In the next two sections, we present different models $\mathcal{M}$ we use as estimator for each risk target. We first present the dataset we use and then we explain the different models considered.

92.4.1 Data

We apply the proposed strategy on BTC. We collected 36,193 close prices from Poloniex exchange, using its API, from 2016-01-01 to 2020-02-16 on an hourly basis. We look at intraday frequency, since it is not rare to observe severe loss at such frequency, due to a relatively higher volatility in the cryptocurrency market compared to traditional assets.

We split the dataset into two sets, train, from 2016-01-01 to 2019-01-01 00:00:00, and test sets, from 2019-01-01 01:00:00 to 2020-02-16 22:00:00. We keep a relatively large test set in order to produce a final robust estimation of the out-of-sample trading strategy performance with a large backtesting period of 9486 observations.

92.4.2 Neural Networks

We first use two different neural network architectures for the trading signals prediction, the multilayer perceptron (MLP) and the long short-term memory neural network (LSTM) architecture from Hochreiter and Schmidhuber (1997). Neural networks are nonparametric, which allows us to avoid making strong assumptions on the data that are not met in reality and that is often the case for financial time series, for example, with the normality assumption of financial returns. On top, neural networks are universal approximators (Leshno et al. 1993) which means that, in theory, neural networks can arbitrarily closely approximate the true process of r_t . In particular, LSTM neural networks are state-of-the-art for many applications such as speech recognition, text extraction, translation, or handwriting recognition, since plain recurrent neural networks are not suited to nonstationary time series modeling. In finance, deep learning has been used in various researches, in particular by Franke (1999) and Zhang et al. (2020) for portfolio management or by Kim and Won (2018) for volatility forecasting with LSTM in comparison with GARCH. Nevertheless, one has yet to prove the superiority of neural networks compared to simpler parametric models in their application.

92.4.2.1 Trader's Function Mapping

Output In order to train the neural networks, we need to build the training input-output pairs. We could directly use the trading signal s_t from (3) as output labels, nevertheless, since $\text{TVaR}_t^{\alpha,w}$ belongs to the tail of the loss series, s_t suffers from severe class imbalance since $\frac{\sum_{t=w+1}^T \mathbf{1}_{-r_t > \text{TVaR}_t^{\alpha,w}}}{T} \ll 0.5$, which is difficult to handle for machine learning models (Krawczyk 2016). On top, financial returns follow an

asymmetric distribution and suffers from the leverage effect (Black 1976; Christie 1982), commonly defined as volatility rising more rapidly when returns are negative than positive. To address that problem, we introduce a new category for r_t , when r_t belongs to the right tail of the distribution, which allows us to control the misclassification cost of the final model. Indeed, let us introduce the following output variable defined as:

$$Y_t = \begin{cases} 1, & \text{if } -r_t < \text{TVaR}_t^{1-\alpha, w} \\ 2, & \text{if } -r_t > \text{TVaR}_t^{\alpha, w} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

We can define the predicted strategy signals as a function of the predicted output variable $\hat{Y}$ as follows:

$$\hat{s}_t = \begin{cases} 1, & \text{if } \hat{Y}_t = 2 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

and we easily derive the unconditional misclassification costs matrix in terms of excess returns.

Outcome	Prediction		
	0	1	2
0	0	0	$\bar{r}_0 \bar{p}_0$
1	0	0	$\bar{r}_1 \bar{p}_1$
2	0	0	$\bar{r}_2 \bar{p}_2$

where $\bar{r}_i$ is the expected return on class i estimated with $\frac{1}{\sum_{t=w+1}^T \mathbf{1}_{Y_t=i}} \sum_{t=w+1}^T r_t \mathbf{1}_{Y_t=i}$

and $\bar{p}_i$ is the weight of class i defined as $\frac{\sum_{t=w+1}^T \mathbf{1}_{Y_t=i}}{T-w}$ which is the empirical estimate of the unconditional probability $P(Y_t = i)$.

As expected, when using the risk target defined in (4), we can see in Table 1 that one has negative costs for the correct classification of class 2 since $\bar{r}_2 \leq 0$ and by definition, the conditional cost for predicting class 2 instead of 1 is much higher than the one for predicting class 2 instead of 0, since in the former, we miss the opportunity to invest during an extremely positive return period. Thus, we clearly see that the classifier must correctly predict the sign of the tail events in order to maximize the strategy return. In order to do so, we use a 3 neurons output layer with softmax activation function corresponding to the output variable Y_t defined in (8).

Input Layer As we are interested in understanding whether neural networks can, by themselves, learn valuable features to produce trading signals and feature selection is not our interest in this work, we use simple transformations of the returns, which can be seen as momentum features.

Table 1 Class 2 misclassification costs for different TVaR^{α,w}

α	w (hours)	Prediction		
		0	1	2
0.01	24	0.01	0.15	−0.15
	2880	0.01	0.14	−0.14
	4320	0.01	0.13	−0.14
0.025	24	0.01	0.15	−0.15
	2880	0.01	0.14	−0.14
	4320	0.01	0.13	−0.14
0.05	24	0.01	0.15	−0.15
	2880	0.01	0.14	−0.14
	4320	0.01	0.13	−0.14
0.1	24	0.01	0.15	−0.15
	2880	0.01	0.14	−0.14
	4320	0.01	0.13	−0.14

In particular, since price time series are not stationary, we use multiple returns instead, $X_t^p = r_{t-p}$ for different periods p , where $r_{t-p} = p/p_{t-1-p} - 1$ and $p \in \{0, 1, 2, 4, 6, 13\}$. We also use the normalized difference between the returns and the two class thresholds which should help identifying a risk-buildup situation, when the returns either explodes toward the upper class threshold $\text{TVaR}_t^{1-\alpha,w}$ or severely drops toward the lower threshold $\text{TVaR}_t^{\alpha,w}$, which are defined respectively as $U_t^{\alpha,w} = \frac{r_t - \text{TVaR}_t^{\alpha,w}}{\text{TVaR}_t^{\alpha,w}}$ and $D_t^{\alpha,w} = \frac{r_t - \text{TVaR}_t^{1-\alpha,w}}{\text{TVaR}_t^{1-\alpha,w}}$.

For the MLP, we then use the vector $X_t \in \mathbb{R}^8 = (X_t^0, X_t^1, X_t^2, X_t^4, X_t^6, X_t^{13}, D_t^{\alpha,w}, U_t^{\alpha,w})$ as input for each risk target. For the LSTM model, since it is a recurrent neural network, we take advantage of its ability to directly modelize sequential data and we use the same features as above on a certain historical window of length 24, using one historical day to make a prediction for the next hour. Thus, the features become: $X_t^p = (r_{t-p-23}, \dots, r_{t-p})$ for $p \in \{0, 1, 2, 4, 6, 13\}$ and $U_t^{\alpha,w} = \left(\frac{r_{t-23} - \text{TVaR}_t^{\alpha,w}}{\text{TVaR}_t^{\alpha,w}}, \dots, \frac{r_t - \text{TVaR}_t^{\alpha,w}}{\text{TVaR}_t^{\alpha,w}} \right)$ and $D_t^{\alpha,w} = \left(\frac{r_{t-23} - \text{TVaR}_t^{1-\alpha,w}}{\text{TVaR}_t^{1-\alpha,w}}, \dots, \frac{r_t - \text{TVaR}_t^{1-\alpha,w}}{\text{TVaR}_t^{1-\alpha,w}} \right)$. We use the same input vector as for the MLP, but now X_t is in $\mathbb{R}^{24, 8}$. We explain the hidden layer architecture in the next section.

Finally, we train our classifier, using Keras python library, on the training data $D = \{(X_w + 1, Y_w + 1), \dots, (X_T, Y_T)\}$ with Adam algorithm, which is a stochastic gradient descent method based on adaptive estimation of first-order and second-order moments of the gradient (Kingma and Ba 2015) and 128 batch size.

92.4.3 Model Selection

92.4.3.1 Hidden Layers Architecture

On the train set, we performed 10 folds cross-validation for time series, described in Fig. 3, with one fold corresponding to 1 month, in order to test different hidden layers architectures for the MLP model. In this work, we keep the hidden layers relatively simple in order to show how a single architecture can effectively extract features for different risk target output. Based on the cross-validation performance evaluated with AUC score (area under the ROC curve), we choose the model with highest AUC score between class 1 and 2, while still controlling for robustness. That is we select the model with highest median AUC score between class 1 and class 2. After a small tuning, we keep three hidden layers and three fully connected layers with 16, 4, and 2 neurons, respectively, as in Fig. 2.

As regularization technique, we use the state-of-the-art Dropout layer after the first and the second hidden layers with a dropout rate of 0.2, in order to prevent overfitting (Srivastava et al. 2014).

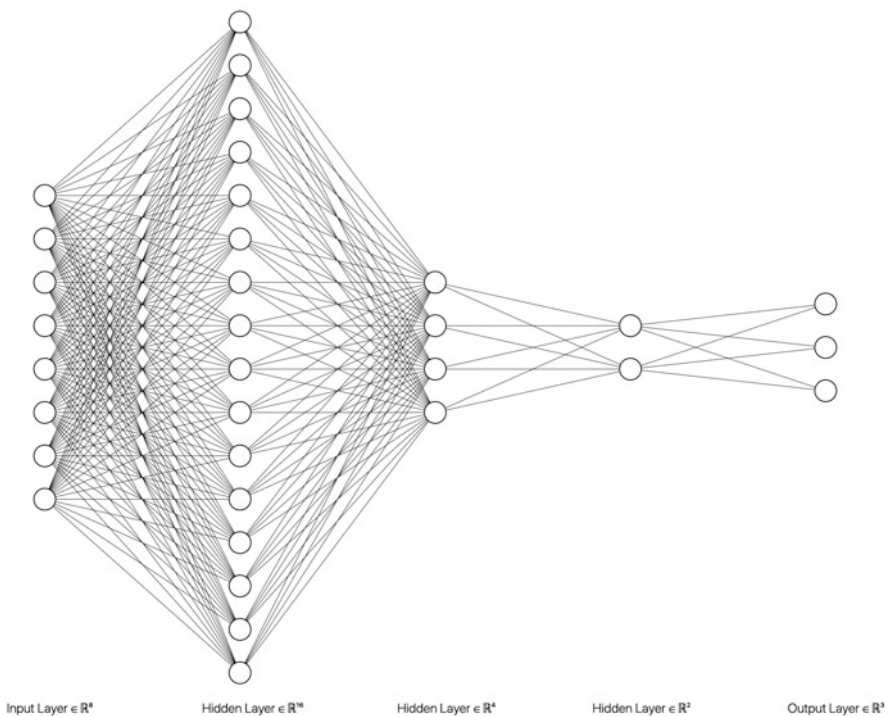


Fig. 2 Final MLP architecture

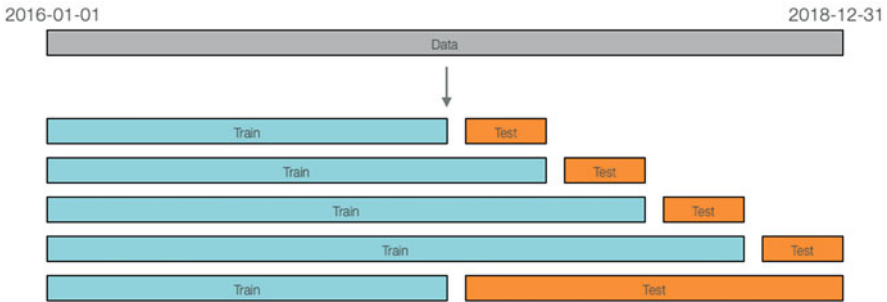


Fig. 3 Time series cross-validation with four folds

As for the LSTM architectures, we simply change the first two fully connected layers to LSTM units in order to test whether LSTM can directly improve the performance of the MLP models. We use the hyperbolic tangent as activation function for the hidden layers.

92.4.3.2 Strategy Parameters and Final Decision

Any classifier $\mathcal{M}$ trained on the dataset D gives for output the probability class p_t . Thus, the trading strategy based on $\mathcal{M}$ has one parameter, the probability threshold u from which we can make a trading decision with respect to (6). The optimal parameter must solve the investor’s goal define in (1) with respect to their risk target. We can now reformulate the optimization problem as it follows (Fig. 3):

$$\begin{aligned} \max_{u,w} \quad & \sum_{t=w+1}^T (R_t - r_t) = \sum_{t=w+1}^T \{(1 - \hat{s}_t)r_t\} \\ \text{s.t.} \quad & \text{VaR}_t^\alpha \leq \text{TVaR}_t^\alpha \end{aligned} \tag{10}$$

Probability Threshold u Denote $\mathcal{U}_\alpha$ the set of u which are tail loss optimal, that is the set of u respecting the constraint formulated in (2). Let us define our strategy with the trading signals $\{\hat{s}_{tw+1}, \dots, \hat{s}_{tT}\}$ and corresponding return $S_{w,\alpha} = \{R_{w+1}, \dots, R_T\}$ for a predefined risk target with level α . We wish to define:

$$\mathcal{U}_\alpha = \{u \in \mathbb{R}, \text{ P}(R_t \leq \text{TVaR}_t^\alpha) \geq 1 - \alpha\}$$

Let us introduce the conditional probability classes for an arbitrary threshold $u \in \mathbb{R}$:

		Prediction	
		0/Hold	1/Sell
Outcome	0/Hold	$\text{TNR}(u) = \text{P}(\hat{s}_t = 0 s_t = 0)$	$\text{FPR}(u) = \text{P}(\hat{s}_t = 1 s_t = 0)$
	1/Sell	$\text{FNR}(u) = \text{P}(\hat{s}_t = 0 s_t = 1)$	$\text{TPR}(u) = \text{P}(\hat{s}_t = 1 s_t = 1)$

where $\text{TNr}(u)$, $\text{FPR}(u)$, $\text{FNR}(u)$, and $\text{TPR}(u)$ stand for *True Negative*, *False Positive*, *False Negative*, and *True Positive Rates* for the threshold u , respectively. In statistical tests, $\text{FPR}(u)$ and $\text{FNR}(u)$ are often called type I and type II errors, respectively. The optimal threshold is given by the optimal trade-off between FPR and FNR , but in order to respect the risk target, one cannot choose any trade-off, since, as one can easily see, $\text{FNR}(u)$ corresponds to the exceedance probability of the portfolio over the risk limit. In practice, as it is pointed out in Packham et al. (2017), such a strategy cannot avoid tail events that occurs “totally out of the blue” (type II error in our formulation). Also, given the empirical stylized fact that return data feature little or no autocorrelation implies that such a strategy may signal a sell order, when ex-post holding the position would have been optimal (type I error).

We can easily see that any threshold $u^* \in \mathcal{U}_\alpha = \left\{ u \in \mathbb{R}, \text{TPR}(u) \geq \frac{\text{P}(-r_t > \text{TVaR}_t^\alpha) - \alpha}{\text{P}(-r_t > \text{TVaR}_t^\alpha)} \right\}$ is tail loss optimal.

Proof We need to compute the following probability for all $u \in \mathbb{R}$ and $t \in [w, T]$:

$$\text{P}(R_t \leq \text{TVaR}_t^\alpha) = 1 - \text{P}(R_t > \text{TVaR}_t^\alpha)$$

We have:

$$\begin{aligned} \text{P}(R_t > \text{TVaR}_t^\alpha) &= \text{P}(R_t > \text{TVaR}_t^\alpha \cap s_t = 1) + \text{P}(R_t > \text{TVaR}_t^\alpha \cap s_t = 0) \\ &= \text{P}(R_t > \text{TVaR}_t^\alpha \mid s_t = 1) \text{P}(s_t = 1) + \underbrace{\text{P}(R_t > \text{TVaR}_t^\alpha \mid s_t = 0) \text{P}(s_t = 0)}_{=0} \\ &= \text{P}(R_t > \text{TVaR}_t^\alpha \mid s_t = 1) \underbrace{\text{P}(s_t = 1)}_{= \text{P}(-r_t > \text{TVaR}_t^\alpha)} \end{aligned} \quad (11)$$

and:

$$\begin{aligned} \text{P}(R_t > \text{TVaR}_t^\alpha \mid s_t = 1) &= \text{P}(R_t > \text{TVaR}_t^\alpha \cap \hat{s}_t = 0 \mid s_t = 1) \\ &\quad + \text{P}(R_t > \text{TVaR}_t^\alpha \cap \hat{s}_t = 1 \mid s_t = 1) \quad \text{P}(\hat{s}_t = 0 \mid s_t = 1) \\ &= \underbrace{\text{P}(R_t > \text{TVaR}_t^\alpha \mid \hat{s}_t = 0 \cap s_t = 1)}_{=0} \quad \text{P}(\hat{s}_t = 0 \mid s_t = 1) \\ &\quad + \underbrace{\text{P}(R_t > \text{TVaR}_t^\alpha \mid \hat{s}_t = 1 \cap s_t = 1)}_{=0} \text{P}(\hat{s}_t = 1 \mid s_t = 1) \\ &= \text{P}(\hat{s}_t = 0 \mid s_t = 1) = \text{FNR}(u) = 1 - \text{TPR}(u) \end{aligned} \quad (12)$$

By plugging (12) in (11), we have:

$$\text{P}(R_t > \text{TVaR}_t^\alpha) = \{1 - \text{TPR}(u)\} \text{P}(-r_t > \text{TVaR}_t^\alpha)$$

Hence:

$$\begin{aligned} P(R_t \leq \text{TVaR}_t^\alpha) &= 1 - \{1 - \text{TPR}(u)\}P(-r_t > \text{TVaR}_t^\alpha) \\ &= 1 - P(-r_t > \text{TVaR}_t^\alpha) + P(-r_t > \text{TVaR}_t^\alpha)\text{TPR}(u) \end{aligned}$$

In order to achieve the goal of our tail risk protection strategy, we must have:

$$\begin{aligned} P(R_t \leq \text{VaR}_t^{\alpha, \text{target}}) &\geq 1 - \alpha \\ 1 - P(-r_t > \text{TVaR}_t^\alpha) + P(-r_t > \text{TVaR}_t^\alpha)\text{TPR}(u) &\geq 1 - \alpha \\ \text{TPR}(u) &\geq \frac{P(-r_t > \text{TVaR}_t^\alpha) - \alpha}{P(-r_t > \text{TVaR}_t^\alpha)} \end{aligned} \quad (13)$$

Thus, any threshold u^* with satisfies (13) is tail loss optimal, since it will ensure a maximum target VaR by controlling the type II error.

Finally, to decide for the final probability class threshold u^* , we simply choose the threshold that maximize the cumulative excess return on the train set and apply it on the corresponding test set.

For illustration, in Fig. 4, we plotted the shape of the constraint on TPR with respect to the risk limit of the investor for different risk level α . As we can see, as the target risk for a level α diverges from the α -quantile of the true distribution of the returns, the constraint on the classifier performance strengthens.

As described in Sect. 92.3, the risk targets are defined by Eq. (4). In Table 2, we show the different constraints from Eq. (2) on the classifier in order to achieve the goal of the tail loss strategy based on the hourly returns of btc on the train set.

As we can see, using a small rolling window to compute the risk target is more conservative; thus, the classifier must have a large TPR in order to ensure a maximum risk target of the final strategy.

92.5 Auto-Regressive Models

Since neural networks are complex nonlinear multifactor models, it is justified to ask if a simple parametric model calibrated on the same training set can catch tail-events as good ML methods or better. In order to compare both approaches, we used three different parametric and a LPA technique from Spokoyny (009a), that we present in this section.

92.5.1 ARMA-GARCH Type Models

92.5.1.1 ARMA-GARCH

As a benchmark, we first used an ARMA-GARCH model for BTC loss series. Indeed, GARCH, developed by Bollerslev (1986), catches well volatility clusters,

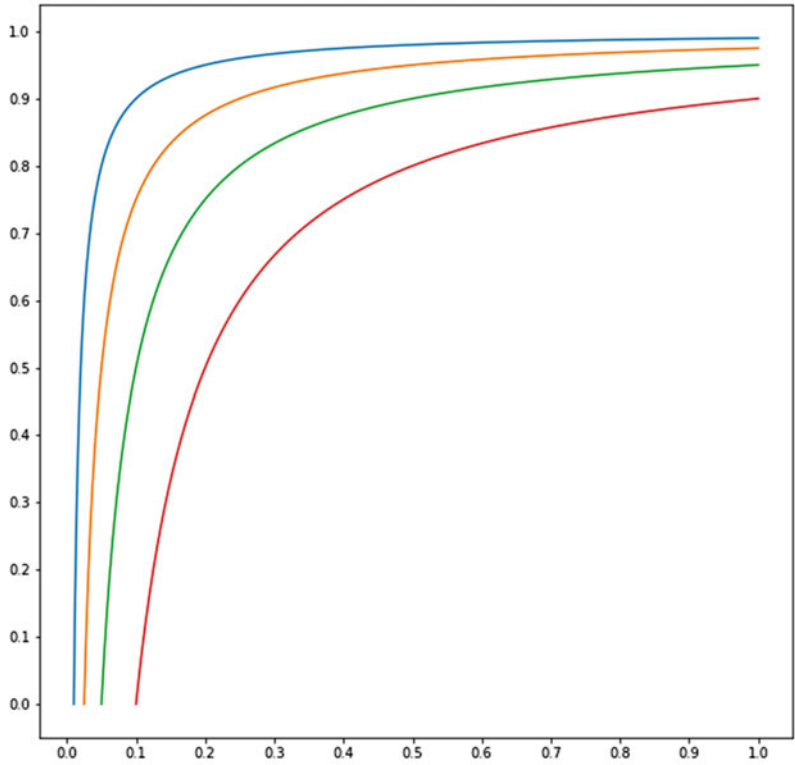


Fig. 4 Expected exceedance and corresponding minimum TPR for significance levels 0.01, 0.025, 0.05, 0.1

Table 2 Expected exceedance and corresponding minimum TPR for different VaR estimators with parameters (α, w), $w = 2880$ and $w = 4320$ correspond to 4 months and 6 months respectively

α	w (hours)	Exceedance	min TPR (%)
0.01	24	4.5	78
	2880	1.2	18
	4320	1.1	11
0.025	24	5.5	55
	2880	2.7	7
	4320	2.6	5
0.05	24	8.0	38
	2880	5.2	4
	4320	5.1	3
0.1	24	12.5	20
	2880	9.7	0
	4320	9.4	0

which often appears in returns time series and ARMA model allows for a time-varying mean of the returns. The ARMA(P, Q)-GARCH(q, p) model is defined as:

$$r_t = a_1 r_{t-1} + a_2 r_{t-2} + \dots + a_P r_{t-P} \\ + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_Q \varepsilon_{t-Q}$$

with

$$\varepsilon_t = Z_t \sigma_t \\ Z_t \sim N(0,1) \\ \sigma_t^2 = \omega + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2$$

where $\omega > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$ and $\sum_{i=1}^p \beta_i + \sum_{j=1}^q \alpha_j < 1$.

92.5.1.2 ARMA-EVTGARCH

Assuming normally distributed innovations, Z , is often not justified and, since in our case we are interested in modeling extreme events, we also fitted an ARMA-GARCH model with a Generalised Pareto distribution (GPD) for the innovations of the GARCH process. This approach has been studied numerous time in the literature, in particular in Packham et al. (2017) or Taylor and Yu (2016) for risk management in trading. We refer to that model as ARMA-EVTGARCH.

The distribution function of the GPD is given by:

$$G_{\xi, \beta}(x) = \begin{cases} 1 - (1 + \xi x / \beta)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp^{-x/\beta}, & \xi = 0 \end{cases}.$$

where $\beta > 0$, $x \geq 0$ when $x \geq 0$ and $0 < x \leq -\beta/\xi$ when $\xi < 0$.

The GPD describes the tail of the data since only the extremes outcomes are included in the estimation with the method of threshold exceedances as we explain in the next paragraph.

92.5.1.3 Model Selection, Calibration, and Exceedance Probability

To find the proper lag orders, P and Q , for the mean, and p and q , for the volatility processes, we use the classical Box-Jenkins method and AIC criterion, see Chen et al. (2018) for a detailed explanation. We perform the model selection on the last 2280 observations (4 months) of the train set. The final model selected is an ARMA(3, 1)-GARCH(1, 2); we use the same orders ARMA-EVTGARCH.

We calibrate the parameters of ARMA-GARCH and ARMA-EVTGARCH simultaneously via quasi-maximum likelihood estimation (QMLE) (McNeil and Frey 2000) using a rolling window of fixed length of 4 months (2880 observations). The size of the rolling window is relatively large for GARCH modeling, nevertheless, in this chapter, we are using intraday data which justifies that choice. As a comparison Taylor and Yu (2016) used a rolling window of 2500 observations. The

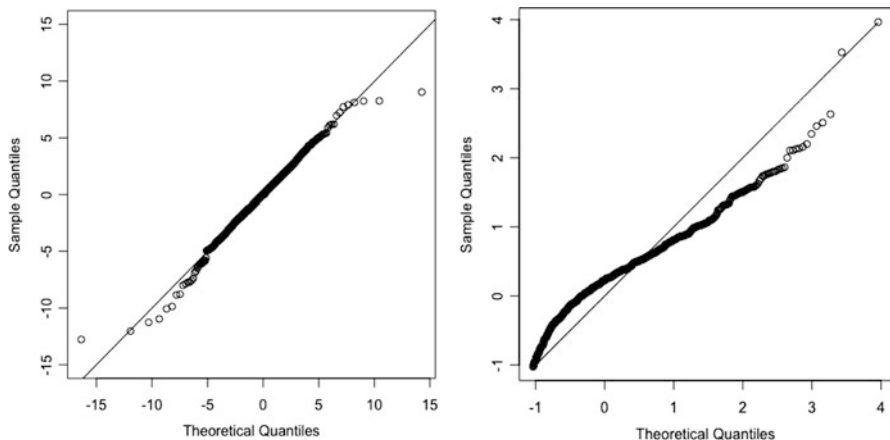


Fig. 5 QQ plots of GARCH residuals. Left: normal distribution; right: GPD distribution (samples beyond upper 5% threshold, g)

calibration of the GPD is based on the upper 5% of the residuals of the ARMA-GARCH; we denote the upper 5% threshold as g . For illustration, residual QQ-plots based on hourly BTC loss data are given in Fig. 5.

The estimated probability class of for the category 2 of Y_t , that is, $P(-r_t > \text{TVaR}_t^\alpha)$, is then given by $p_t = 1 - F((\text{TVaR}_t^\alpha - \hat{\mu}_{t+1})/\hat{\sigma}_{t+1})$ where $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ are the models' mean and volatility forecasts and F is the normal distribution function for the ARMA-GARCH and GPD for the ARMA-EVTGARCH model. For computing p_t of ARMA-EVTGARCH, the GPD is defined only for $x \geq g$, so if $(\text{TVaR}_t^\alpha - \hat{\mu}_{t+1})/\hat{\sigma}_{t+1} < g$, F is the normal distribution, since it is not necessary to use a GPD distribution for observations that do not belong to the tail. We recalibrate the parameters of the models every hour for each new observation in the test set.

92.5.2 CARL-vol Model

The benefit of ARMA-GARCH type models is their effectiveness at catching stylized facts of financial time series and their simplicity, nonetheless, they do not directly model the exceedance probability p_t , but the total distribution of the return r_t . In order to use the advantages of GARCH models and to directly modelize the exceedance probability, we also fit a CARL-vol model from Taylor and Yu (2016) which specifies p_t for a constant threshold $Q \in \mathbb{R}$ as:

$$p_t = \frac{0.5}{1 + \exp(-x_t)} + 0.5\mathbf{I}(Q > 0)$$

where $x_t = \phi_0 + \phi_1\sigma_t$ and we assume a GARCH(1, 1) for the return, that is, $\sigma_t^2 = \omega + \beta_1\sigma_{t-1}^2 + \alpha_1(r_{t-1} - \mu)^2$. ϕ_i , α_i , β_i , and μ are constant parameters.

We calibrate the parameters of CARL-vol with Maximum Likelihood Estimation using a Bernoulli distribution with density specified as $f(r_t) = p_t^{\mathbf{1}_{(y_t \leq Q)}} (1 - p_t)^{1 - \mathbf{1}_{(y_t \leq Q)}}$. For the model calibration, we use the same rolling window as for the GARCH-based models of the previous section, where Q is set to TVaRt of the last observation of the rolling window. Since CARL-vol assumes a fixed threshold Q and that, in our case, it is time-varying, we also recalibrate the model every hour, estimating new parameters for each observation in an online manner.

92.5.3 Local Parametric Approach (LPA)

For the calibration of the three previous models, we use a fixed rolling window, assuming time homogeneity on that interval, in the sense that the process r_t follows the same structural equation at each time point defined by ARMA-GARCH, ARMA-EVTGARCH, and CARL-vol models. This approach, of course, does not take into consideration structural breaks (Bouri et al. 2019). Ignoring breaks will increase the bias (Hillebrand 2005). Though, in order to relax time homogeneity, we can either use a nonparametric model which allows for time-varying parameters, as we do with the neural network models presented in chapter 92.4, or we can use a Local Parametric Approach (LPA) thanks to the Local Change Point detection method suggested by Spokoiny (1998) and developed by Härdle et al. (2003), Mercurio and Spokoiny (2004) and Spokoiny (2009) in the context of volatility modeling.

As in Cizek et al. (2009), we assume GARCH(1, 1), characterized by parameters θ , is a good candidate for the process r_t on a local scale. For estimating θ , we apply the QMLE approach using a Gaussian distribution for the innovations. The LPA allows us to find the largest interval of homogeneity, $\mathcal{I}(t)$, that is, the longest interval $\mathcal{I}$ with the right-end point t , where data do not contradict the assumption that θ is constant. To find the change points, we use the test procedure proposed by Klochkov et al. (2019) that uses the multiplier bootstrap method from Spokoiny and Zhilova (2015) to find the critical values for the likelihood ratio test, generating 100 bootstrap weights sequences for the test. The computational cost is high and our test set is quite large; thus, in practice, we choose a fixed decreasing size of 5 observations for the successive intervals tested and we only performed the LPA detection every 5 observations in the test set. This does not come without cost and automatically increases the bias of $\hat{\theta}$, nevertheless, it decreases the bias in comparison of a fixed window size approach, as for the GARCH-based models of the previous sections. We denote the model calibrated with LPA method as LCP-GARCH.

92.6 Applications

92.6.1 Meta Strategy: Stacking

In the previous section, we presented six different classifiers to build trading signals for tail events hedging. Since the GARCH-based approach is essentially linear, but MLP and LSTM are nonlinear classifiers, we expect them to extract different

features from the data. A direct question is then to know which model to use in which situation. To address that problem, we propose to combine all models output from the test set into a meta model with a stacking classifier. Stacking is an ensemble learning technique where the predictions of multiple classifiers (referred as level-0 classifiers) are used as new features to train a meta-classifier (level-1 classifier).

The meta-classifier is trained on the predictions made by level-0 classifiers on the out-of-sample data. That is, data which has not been used to train the level-0-classifiers is fed to them in order to get the predictions $p_t = (p_t^j)$, where $1 \leq j \leq 6$ corresponds to the level-0 models and t belongs to the test set indices. We then use the predictions to build input features and the corresponding labels Y_t to build the training input-output pairs for the meta classifier.

Our meta-classifier should be explainable in order to directly interpret which level-one classifier contributes the most to the final meta-prediction. For its simplicity and interpretability, we use logistic regression. We have for all $t \in [T - n - 1, T]$ where n is the test set size:

$$s_t = \sum_{j=1}^6 \beta^j P_t^j$$

where $P_t^j = \log \frac{p_t^j}{1-p_t^j}$ and p_t^j is the output of each one-level classifier that is the estimated probability of class 2 of Y_t . Moreover, since we need to interpret the coefficients β^j , we rescale p_t^j to $[0, 1]$ range.

When using stacking technique, the issue is the multicollinearity of the prediction of the level 1 classifiers. In particular, we can expect ARMA-GARCH and ARMA-EVTGARCH models to present multicollinearity in their output. To address that problem, we use L2 regularization to improve robustness, performing a Ridge type logistic regression (Hoerl and Kennard 1970), where the final estimator for the model coefficients is the solution of:

$$\min_{\beta_1, \dots, \beta_6} \sum_{i=T-n-1}^T \left(s_t - \sum_{j=1}^6 \beta^j p_t^j \right)^2 + \lambda \sum_{j=1}^6 \beta^j{}^2$$

where $\lambda \in \mathbb{R}^+$ is the regularization hyper-parameter. In practice, we must finetune this parameter by performing, for example, cross-validation. In our setup, we did not tune it since cross-validation would be too computationally intensive and fixed it to $\lambda = 1$. We will denote that model as Ensemble.

In the next sections, we present our results. They correspond to out-of-sample performance of the different models on the test set from 2019-01-01 to 2020-02-16. The neural network models are retrained every 824 observations corresponding to the size of the validation folds in order to reflect the validation performance, whereas the ARMA-GARCH, ARMA-EVTGARCH, CARL-vol, and Ensemble are recalibrated every new observation and LPA-GARCH every five observations as stated in the previous section.

Every hour, each model generates the trading decision $\hat{s}_t$ based on the data available until t , then it is applied to the return, r_{t+1} , on the next period $[t, t + 1]$. The current set-up does not account for market frictions, such as bid-ask spreads, and other associated costs, but we included the trading fees from Binance exchange, which are, at most, 0.1% on each trade. Nonetheless, the cost associated to the fees should be very low since we only trade when a tail event is predicted.

92.6.2 Benchmarks

There are a few well-documented tail risk protection strategies in the literature, such as the constant proportion portfolio insurance (CPPI) strategy (Black and Jones 1987), the Dynamic Proportion Portfolio Insurance (DPPI) or Time-varying Proportion Portfolio Insurance (TPPI) strategy (Hamidi et al. 2009; Happersberger et al. 2019), or the protective put strategy using options. In this chapter, we compared our results with more recent and machine learning oriented work, in particular the Varspread and target VaR strategies. Finally, the last benchmark used is the simple buy-and-hold strategy, where an investor buys BTC at 2019-01-01 00:00:00 and sells it at the end of the test set at 2020-02-16 22:00:00.s.

92.6.3 Out-of-Sample Performance

92.6.3.1 Model Risk Evaluation

In order to assess the performance of each strategy, we first look at the model risk which comes from the statistical performance of the classifier. A familiar metric is the Area Under the ROC Curve (AUC), nevertheless, the classical AUC, while taking into consideration class imbalance, does not include the cost associated with each class. For our trading application, as we stated in Sect. 92.4.2, we have strong cost imbalance between the outcomes. Thus, for our evaluation, we use the risk-adjusted or cost-adjusted AUC by Jordà and Taylor (2011).

The ROC curve is defined for all threshold u as:

$$\begin{aligned} \text{ROC} : [0, 1] &\rightarrow [0, 1] \\ \text{FPR}(u) &\mapsto \text{TPR}(u) \end{aligned}$$

We incorporate the average costs associated with false positives ($c_F = \bar{r}_0 p_0 + \bar{r}_1 p_1$) and true positives ($c_T = -\bar{r}_2 p_2$) in order to get the risk-adjusted curve, denoted AROC and rescaled to $[0, 1]$, defined for all threshold u as:

$$\begin{aligned} \text{AROC} : [0, 1] &\rightarrow [0, 1] \\ c_F \text{FPR}(u) &\mapsto c_T \text{TPR}(u) \end{aligned}$$

The risk-adjusted AUC, denoted AAUC is then $\text{AAUC} = \int_0^1 \text{AROC}(u) du$.

Table 3 Risk-adjusted AUC on the test set

α (%)	w (hours)	MLP	LSTM	ARMA-GARCH	ARMA-EVTGARCH	LPA-GARCH	CARL-vol	Ensemble
1.0	24	69	69	73	73	71	71	73
	2880	62	67	69	72	72	69	74
	4320	70	69	70	72	72	74	74
2.5	24	66	69	70	70	69	69	71
	2880	72	71	74	74	76	72	77
	4320	71	73	74	74	75	73	77
5.0	24	65	65	66	66	65	66	67
	2880	71	70	74	74	75	73	76
	4320	70	71	74	73	75	73	76
10	24	63	63	63	63	60	62	64
	2880	67	66	70	70	71	71	72
	4320	66	68	69	70	70	71	71

In Table 3 we show the risk-adjusted AUC score on the whole test period. We can see that the MLP and LSTM models have similar performance; we can make the same observation for the ARMA-GARCH, ARMA-EVTGARCH, and LPA-GARCH approach model. Nonetheless, we observe that LPA-GARCH performs better than ARMA-GARCH and ARMA-EVTGARCH for stable risk target where $w = 2880$ or $w = 4320$. This indicates that using a model with less complexity fitted on a time-homogeneous window generalized better than more complex models fitted on large window. Finally, all parametric or local parametric models perform better than the neural network-based models in terms of risk-adjusted classification, which is a major drawback to the deep learning approach in our setting.

In Table 4 we present the cross-validation performance on the test set. Indeed recall, that the results from the MLP and the LSTM models are obtained by retraining the models every 824 observations in the same approach as depicted in Fig. 3. For comparison, we also evaluate the other classifiers with the same sampling method.

We can clearly see that ARMA-GARCH and ARMA-EVTARCH outperforms MLP and LSTM classifiers for small significance level and varying risk target ($\alpha = 0.01$ or $w = 0.025$ and $w = 24$) both on average and in terms of robustness. This shows that, in our setup, GARCH-based models still outperforms nonparametric models at catching volatility clusters. On the other hand, for larger significance level and stable risk target ($\alpha = 0.05$ or $\alpha = 0.1$ and $w = 2280$ or $w = 4320$), the nonparametric approaches do at least as good as the simple GARCH-based models on average and in terms of robustness which indicates a higher generalization power.

From the two tables above, we can also conclude that we benefit from combining those classifiers into a stacking classifier, since the AAUC of the Ensemble classifier is always higher or equal to the maximum AAUC of the weak classifiers on the whole test period. Moreover, in terms of cross-validation performance, the Ensemble

Table 4 Average risk-adjusted AUC and (variance risk-adjusted AUC) over the testing periods

α (%)	w (hours)	MLP	LSTM	ARMA-GARCH	ARMA-EVTGARCH	LPC-GARCH	CARL-vol	Ensemble
1.0	24	70 (0.3)	70 (0.4)	74 (0.1)	74 (0.1)	72 (0.2)	72 (0.2)	73 (0.2)
	2880	58 (3.3)	64 (3.9)	60 (0.9)	61 (1.3)	61 (1.4)	55 (1.0)	62 (1.8)
	4320	63 (3.7)	63 (3.1)	58 (1.2)	62 (1.0)	59 (1.3)	57 (0.8)	60 (1.6)
2.5	24	68 (0.4)	70 (0.1)	71 (0.1)	71 (0.1)	69 (0.1)	70 (0.2)	71 (0.1)
	2880	66 (0.9)	65 (1.3)	62 (1.3)	62 (1.2)	64 (1.2)	57 (1.3)	66 (1.1)
	4320	63 (0.6)	65 (2.0)	62 (1.4)	62 (1.6)	61 (1.2)	57 (1.3)	65 (1.6)
5.0	24	67 (0.4)	67 (0.2)	67 (0.1)	67 (0.1)	65 (0.2)	66 (0.2)	68 (0.2)
	2880	65 (0.7)	64 (0.6)	62 (0.8)	62 (1.0)	63 (0.8)	61 (1.1)	64 (1.0)
	4320	66 (0.5)	64 (0.7)	62 (0.9)	61 (1.0)	63 (0.9)	60 (0.8)	64 (1.0)
10	24	64 (0.2)	64 (0.2)	64 (0.1)	64 (0.1)	61 (0.1)	63 (0.2)	65 (0.2)
	2880	61 (0.7)	63 (0.8)	61 (0.8)	61 (0.8)	61 (0.7)	60 (0.7)	62 (0.8)
	4320	62 (0.7)	64 (0.7)	61 (0.7)	61 (0.7)	61 (0.8)	60 (0.6)	62 (0.9)

classifier always outperforms the average performance, except for the strategy with TVAR^{0.01,4320} risk target. In particular, for larger risk level ($\alpha = 0.025$, $\alpha = 0.5$ and $\alpha = 0.1$) and small window ($w = 24$), the Ensemble classifier outperforms the best performing weak classifier. This result is a clear indicator that each classifier can extract information which is uncorrelated with some of the other classifiers for highly adaptive risk target. This is a sign that each classifier might be more adapted to specific market regimes, but that the Ensemble model is more general.

Overall, from those metrics, we can say that the nonparametric models present more risk than the parametric ones or the Ensemble classifier. Now, in the next section, we will relate the model risk to the economic risk via backtest performance of the associated trading strategies.

92.6.3.2 Backtest Performance

Risk Profile First, we must verify that the strategy associated with each classifier is verifying the target risk constraint. For this we compare, in Table 5, the exceedance of each strategy with the risk level α , if the exceedance is larger than α , then the strategy does not respect the investor’s risk preferences.

All classifiers achieve the risk targets, but both benchmark target VaR strategies, where VaR is estimated with ARMA-GARCH and ARMA-EVTGARCH, do not respect the constraint for TVaR^{1%,24}. The Varspread strategy is not a target VaR strategy, but we also look at its exceedance and we see that it also fails for the same risk target, TVaR^{2.5%,24} and TVaR_{5%,24}, indicating difficulties for those strategies to adapt to varying risk preferences.

Now, in Table 6, we compare the average return of the strategies associated with each classifier and the benchmarks on the whole test period. Achieving a larger average return, while having lower exceedances indicates a better risk-profile for a tail risk averse investor. First, the Ensemble strategy always has a larger average

Table 5 Exceedances of tail risk protection strategies in %

α (%)	w (hours)	Buy& Hold	MLP	LSTM	ARMA-GARCH	ARMA-EVTGARCH	LPC-GARCH	CARL-vol	Ensemble	target VaR norm	target VaR evt	Varspread
1.0	24	4.3	0.3	0.4	0.5	0.4	0.9	1.2	1.0	1.4	1.1	3.6
	2880	1.0	0.6	0.7	0.8	0.6	0.9	0.4	0.7	0.5	0.5	0.7
	4320	0.9	0.7	0.5	0.7	0.5	0.8	0.4	0.7	0.4	0.4	0.7
2.5	24	5.2	1.5	1.8	1.2	1.0	2.1	2.0	1.8	1.8	1.9	4.3
	2880	2.5	2.1	1.6	1.6	1.3	1.2	1.2	1.7	1.2	1.3	2.1
	4320	2.2	2.0	1.3	0.9	0.8	0.8	1.1	1.7	1.0	1.0	2.0
5.0	24	7.9	3.6	3.6	2.8	3.1	4.5	3.4	2.5	2.6	3.6	5.9
	2880	4.9	4.0	3.2	2.3	2.4	2.3	2.5	3.5	2.3	2.8	3.4
	4320	4.5	3.8	3.1	1.4	1.5	3.6	2.3	3.3	2.2	2.5	3.3
10	24	12.4	8.4	8.1	4.0	4.0	8.3	6.6	8.8	4.4	7.2	9.7
	2880	9.6	7.9	7.2	5.6	5.6	5.1	5.7	7.5	4.4	6.1	7.2
	4320	9.4	7.4	7.0	3.3	3.2	4.4	6.0	7.1	4.5	5.7	7.2

Table 6 Average return in % (e-3). Buy&Hold: 10e-3%

α (%)	w (hours)	MLP	LSTM	ARMA-GARCH	ARMA-EVTGARCH	LPC-GARCH	CARL-vol	Ensemble	target VaR norm	target VaR evt	Varspread
1.0	24	4.9	5.1	2.3	1.6	6.5	-15.0	8.9	6.5	6.3	4.0
	2880	8.0	5.5	10.0	4.8	8.6	17.0	4.4	7.1	6.3	4.0
	4320	8.1	11.0	8.8	7.0	9.5	12.0	14.0	8.2	7.1	4.0
2.5	24	7.3	7.1	3.4	6.3	8.9	-0.7	6.8	7.0	7.5	3.7
	2880	6.9	9.9	3.6	1.4	2.2	12.0	8.5	6.4	6.3	3.7
	4320	5.8	10.0	5.3	4.5	6.3	12.0	8.3	7.4	7.5	3.7
5.0	24	7.2	9.2	5.8	6.7	9.3	2.9	7.1	7.8	8.3	2.7
	2880	6.8	13.0	0.8	1.4	2.7	13.0	8.7	5.9	6.2	2.7
	4320	8.5	12.0	1.5	2.1	5.1	16.0	11.0	6.5	7.2	2.7
10	24	5.7	9.6	5.5	6.1	2.8	6.9	6.9	5.9	7.8	2.4
	2880	10.0	7.5	1.6	1.8	5.5	14.0	6.8	5.5	6.3	2.4
	4320	14.0	9.8	3.8	4.1	5.5	12.0	8.1	5.8	7.3	2.4

return than the average performance of the weak strategies, that is, the strategies associated with the weak classifiers, except for the $\text{TVaR}^{1\%,2880}$ risk target, and has larger average return than the maximum of the weak strategies for the $\text{TVaR}^{1\%,24}$ and $\text{TVaR}^{1\%,4320}$. In general, the Ensemble strategy also outperforms the benchmark strategies, except for the target VaR strategies with $\text{TVaR}^{1\%,2880}$, $\text{TVaR}^{2.5\%,24}$, $\text{TVaR}^{5\%,24}$, and $\text{TVaR}^{10\%,24}$. Only these strategies are better suited than our offered Ensemble strategy for their respective investor’s profile, as they offer larger average return, while respecting the corresponding risk target. This result shows that traders can benefit from combining classical econometrics and machine learning oriented approaches in order to offer better tail risk protection strategies to their investors.

Historical Performance In Fig. 6, we show the historical performance on the test period and on Fig. 7, the excess return over the BTC Buy&Hold benchmark in strong uptrend period. We can clearly see how Varspread strategy is too conservative in

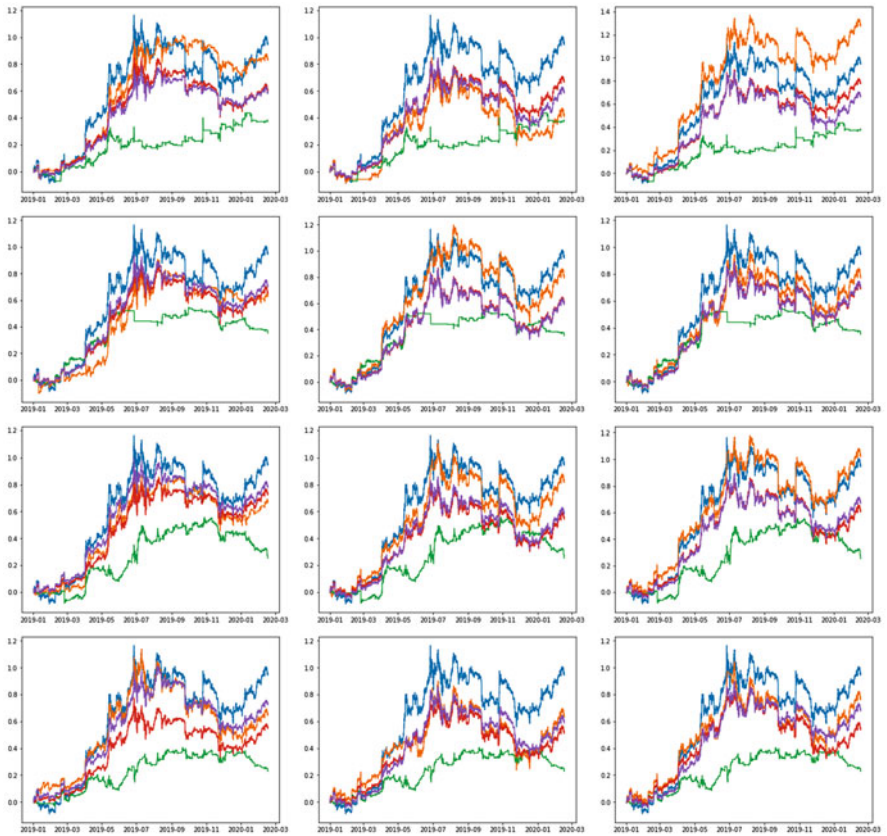


Fig. 6 Ensemble strategy returns on test period for $\text{TVaR}^{\alpha,w}$ with $\alpha \in \{1\%, 2.5\%, 5\%, 10\%\}$ from top to bottom and $w \in \{24, 2880, 4320\}$ from left to right with BTC Buy&Hold, Varspread, target VaR norm, and target VaR evt benchmarks

period of uptrend which severely affects its total return. While the Ensemble strategy significantly outperforms the other actively managed tail risk protection strategies in terms of total return for $\text{TVAR}^{1\%,24}$, $\text{TVAR}^{1\%,4320}$, $\text{TVAR}^{2.5\%,2880}$, $\text{TVAR}^{2.5\%,4320}$, $\text{TVAR}^{5\%,2880}$, and $\text{TVAR}^{5\%,4320}$. For the other risk targets, the historical performance is similar. In the meantime, only two strategies outperform the simple Buy&Hold strategy on the whole test period, for $\text{TVAR}^{2.5\%,4320}$ and $\text{TVAR}^{5\%,4320}$ risk targets, managed by our proposed Ensemble model. This shows that for BTC asset, it is very hard to find the right trade-off between investment during strong uptrend and deinvestment in period of downtrend and that strategies based on simpler models, as Varspread or target VaR strategies, cannot beat the Buy&Hold benchmark in terms of total return, as they overshoots risk in strong uptrend periods.

Indeed, as you can see in Fig. 7, in strong uptrend period, all strategies underperform the Buy&Hold benchmark except for the Ensemble strategy with $\text{TVAR}^{1\%,4320}$ risk target.

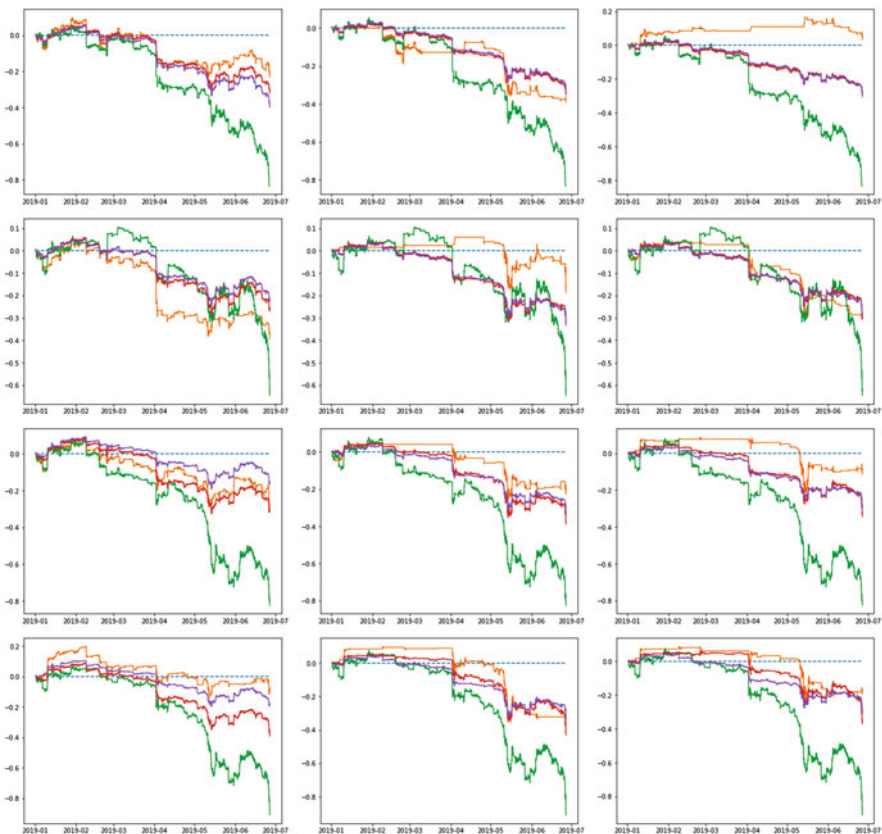


Fig. 7 Ensemble strategy excess returns on uptrend period for $\text{TVAR}^{\alpha,w}$ with $\alpha \in \{1\%, 2.5\%, 5\%, 10\%\}$ from top to bottom and $w \in \{24, 2880, 4320\}$ from left to right with Varspread, target VaR norm and target VaR evt benchmarks

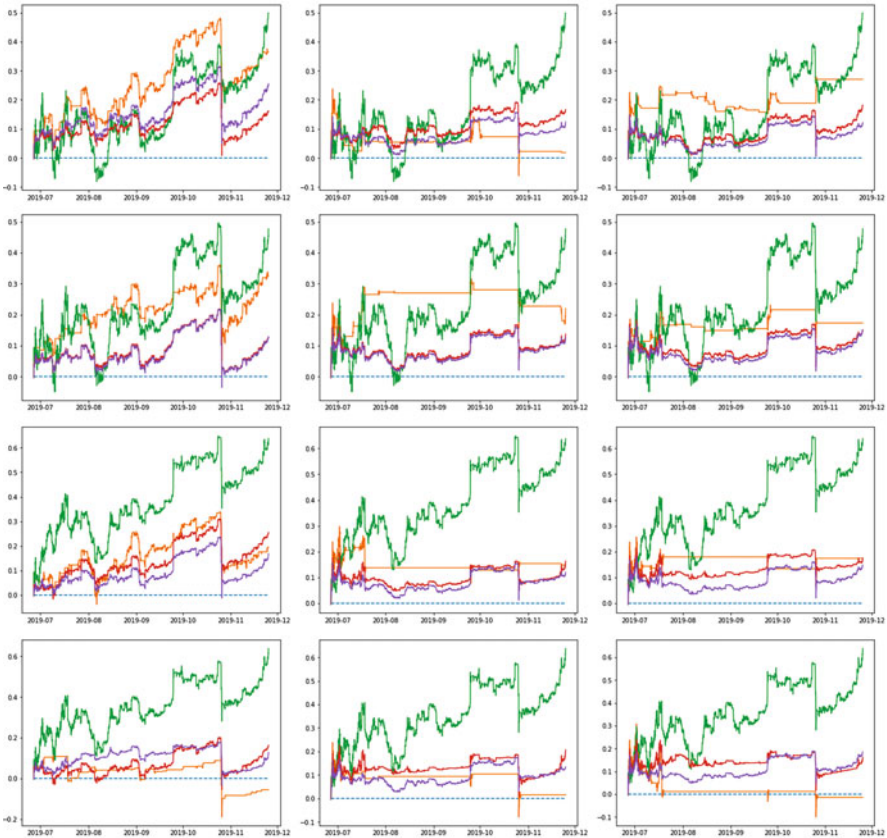


Fig. 8 Ensemble strategy excess returns on downtrend period for $\text{TVAR}^{\alpha,w}$ with $\alpha \in \{1\%, 2.5\%, 5\%, 10\%\}$ from top to bottom and $w \in \{24, 2880, 4320\}$ from left to right with Varspread, target VaR norm, and target VaR evt benchmarks

Nevertheless, as we can see in Fig. 8, that, in period of downtrend from 2019-06-26 to 2019-11-25 where BTC suffered a 60% drawdown, the Varspread strategy is very efficient at catching tail events achieving 50%, 48%, 64%, and 64% excess return for $\alpha \in \{1\%, 2.5\%, 5\%, 10\%\}$, respectively, even generating a positive total return of 4% for $\alpha \in \{5\%, 10\%\}$.

Switching Strategy In order to benefit from the effective tail risk protection from the Varspread strategy in bear market and the well calibrated risk protection of the Ensemble strategy in bull market, we build a final switching strategy which either invest in the Varspread strategy or in the Ensemble strategy depending on the market regime. In order to decide when to switch strategy, we build a simple estimator of the trend by using a simple moving average of BTC price, defined for $n \in \mathbb{N}$ as $\text{MA}_{t-1,n} = \frac{1}{n} \sum_{i=1}^n P_{t-i}$. We define the following trend indicator δ_t as in Rickenberg (2019):

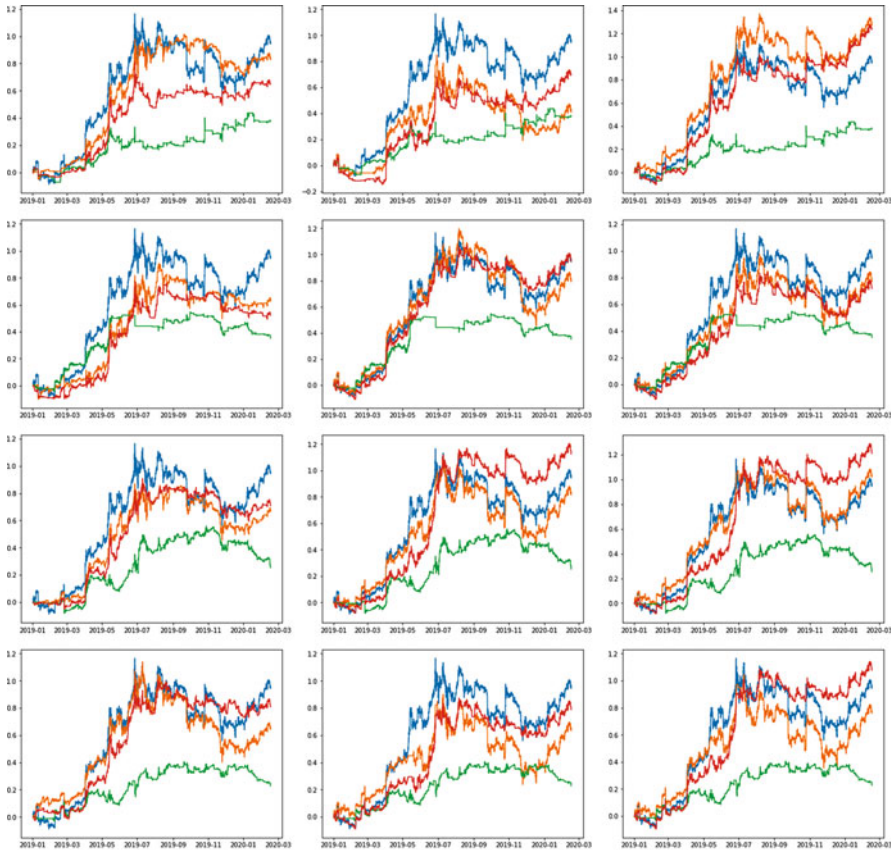


Fig. 9 Historical performance on the test period of Ensemble, Varspread, Switch, and Buy&Hold strategies

$$\delta_t = \begin{cases} 0, & \text{if } P_{t-1} \leq \text{MA}_{t-1,n} \\ 1, & \text{if } P_{t-1} > \text{MA}_{t-1,n} \end{cases}$$

Then we define the weights for each strategy as $w_t^{Ens} = \delta_t$ and $w_t^{Varspread} = 1 - \delta_t$. The historical performance of the final meta-strategy, denoted Switch strategy, is presented in Fig. 9.

In Table 7, we present different backtest statistics for the benchmark, Ensemble, and final Switch strategies. First, we see that the Switch strategy has the highest total return on the testing period for relatively high risk target, but also for $\text{TVAR}^{1\%,2880}$ and $\text{TVAR}^{2.5\%,2880}$. Moreover, we can see that the Switch strategy provides an enhanced risk-return profile indicated by larger Sharpe ratio for all risk targets, than the average between the Varspread and the Ensemble strategies. For example, with $\text{TVaR}^{5\%,2880}$, the Switch strategy increased the Sharpe Ratio by 62% and 219% for the Ensemble and Varspread strategies, respectively. On top, the switch strategy

Table 7 Backtest performance of switch strategies: Total return (%), Sharpe ratio, Sortino ratio, Max Drawdown (MDD), Valut-at-Risk, and Volatility of different strategies on the test period. The total return, Sharpe, and Sortino ratio are annualized. Legend: winner/loser

α (%)	w (hours)	Strategy	Total return	Sharpe ratio	Sortino ratio	MDD	volatility
1.0	24	Buy&Hold	86	1.37	2.62	52	0.69
		Ensemble	76	1.52	2.52	31	0.54
		Switch	59	1.57	2.29	30	0.41
		Target VaR evt	54	1.28	2.13	33	0.46
		Target VaR norm	55	1.24	2.10	40	0.49
		Varspread	34	1.14	1.53	17	0.33
	2880	Ensemble	38	0.70	1.44	53	0.60
		Switch	63	1.50	2.36	27	0.46
		Target VaR evt	54	1.16	2.15	45	0.50
		Target VaR norm	61	1.32	2.37	40	0.50
		Varspread	34	1.14	1.53	17	0.33
	4320	Ensemble	115	1.97	3.91	43	0.64
		Switch	111	2.39	3.95	23	0.51
		Target VaR evt	60	1.31	2.37	43	0.50
		Target VaR norm	71	1.54	2.75	39	0.50
		Varspread	34	1.14	1.53	17	0.33
2.5	24	Ensemble	58	1.14	1.71	36	0.55
		Switch	47	1.27	1.90	27	0.41
		Target VaR evt	64	1.40	2.37	42	0.50
		Target VaR norm	60	1.32	2.21	42	0.50
		Varspread	34	1.35	2.23	18	0.26
	2880	Ensemble	73	1.34	2.71	55	0.59
		Switch	85	2.05	4.15	32	0.46
		Target VaR evt	54	1.19	2.21	45	0.50
		Target VaR norm	55	1.23	2.30	45	0.49
		Varspread	34	1.35	2.23	18	0.26
	4320	Ensemble	71	1.26	2.54	45	0.62
		Switch	65	1.50	2.91	30	0.47
		Target VaR evt	64	1.43	2.55	41	0.49
		Target VaR norm	63	1.42	2.55	41	0.49
		Varspread	32	1.35	2.23	18	0.26
5.0	24	Ensemble	61	1.25	1.86	39	0.53
		Switch	64	1.64	2.69	29	0.43
		Target VaR evt	71	1.58	2.82	39	0.49
		Target VaR norm	66	1.53	2.67	34	0.47

(continued)

Table 7 (continued)

α (%)	w (hours)	Strategy	Total return	Sharpe ratio	Sortino ratio	MDD	volatility
		Buy&Hold	86	1.37	2.62	52	0.69
		Varspread	23	0.70	1.33	27	0.36
		Ensemble	75	1.38	2.60	54	0.59
		Switch	101	2.23	5.08	27	0.50
		Target VaR evt	54	1.18	2.19	45	0.49
		Target VaR norm	50	1.17	2.20	44	0.47
		Varspread	23	0.70	1.33	27	0.36
	2880	Ensemble	92	1.68	2.92	49	0.61
		Switch	108	2.38	5.09	23	0.50
		Target VaR evt	61	1.38	2.48	42	0.49
		Target VaR norm	56	1.29	2.39	42	0.47
		Varspread	23	0.70	1.33	27	0.36
	4320	Ensemble	92	1.68	2.92	49	0.61
		Switch	108	2.38	5.09	23	0.50
		Target VaR evt	61	1.38	2.48	42	0.49
		Target VaR norm	56	1.29	2.39	42	0.47
		Varspread	23	0.70	1.33	27	0.36
10	24	Ensemble	59	1.07	1.94	55	0.61
		Switch	73	1.70	3.03	26	0.47
		Target VaR evt	66	1.43	2.65	42	0.51
		Target VaR norm	51	1.22	2.01	39	0.45
		Varspread	21	0.74	1.18	17	0.30
		Varspread	21	0.74	1.18	17	0.30
	2880	Ensemble	58	1.07	1.95	52	0.60
		Switch	71	1.69	3.33	27	0.46
		Target VaR evt	54	1.21	2.29	44	0.49
		Target VaR norm	47	1.18	2.16	40	0.44
		Varspread	21	0.74	1.18	17	0.30
	4320	Ensemble	70	1.32	2.33	53	0.58
		Switch	97	2.27	4.50	23	0.47
		Target VaR evt	62	1.41	2.57	40	0.48
		Target VaR norm	50	1.25	2.28	44	0.43
		Varspread	21	0.74	1.18	17	0.30

has the best risk-return profile for all risk targets, clearly outperforming the Buy&Hold and other benchmarks, except TVAR^{2.5%,24} where the Varspread strategy has the largest Sharpe ratio. This indicates that the switch strategy is also better suited for Markowitz type optimization where the objective is the mean-variance performance. For the Ensemble strategy itself, the results are mixed with respect to mean-variance performance, since it has a relatively high volatility for all risk targets. Indeed, finding the right trade-off between participation in bull market regimes and hedging tail events comes with a cost in terms of volatility. Since in our case, we know that the Switch and Ensemble strategies are respecting the tail risk constraint, we are interested in the Sortino ratio, which penalizes only volatility

associated with negative returns and is defined as $S = \frac{R}{\sigma^-}$, where R is the expected return and σ^- is the portfolio downside standard deviation. We can see that the Switch strategy clearly outperforms other benchmarks, except for conservative strategies with $\text{TVAR}^{1\%,24}$, $\text{TVAR}^{1\%,2880}$, $\text{TVAR}^{2.5\%,24}$, and $\text{TVAR}^{5\%,24}$.

92.7 Conclusion

We build a dynamic trading strategy in which the goal is to protect against severe losses by targeting a maximum Value-At-Risk defined for any significance level. For this, we model distributional properties of BTC returns in time with nonparametric models such as MLP and LSTM neural networks and compare their performance to local-parametric GARCH, parametric GARCH fitted with various distribution for the innovation, and CARL-Vol models. In our approach, we directly take into account varying investor's risk preferences and optimize the strategy accordingly, allowing us to compare the models for multiple risk targets.

First, we can conclude that for relatively stable risk preferences, parametric and local parametric approaches have a better generalization than nonparametric methods, but for unstable preferences they work better, taking into account time varying risk targets and adapting to sudden changes in the market. On top, we show that, for BTC returns, a general model performing well enough in all market regimes is too hard to find. A tail loss protection strategy based on a single model, while often outperforming a simple Buy&Hold strategy, is still not optimal, either overshooting the risk target in periods of losses or being too conservative in bull market regimes. That is why, we show the benefits of combining multiple models in an Ensemble classifier, which improves generalization with a higher prediction accuracy and economic gains with an enhanced risk-return profile. Indeed, the Ensemble-based strategy combines advantages of classical GARCH tools, which catches better volatility clusters than more complex MLP or LSTM, and deep learning tools which can adapt to time-varying properties of the data distribution. Further, we show that tail loss protection strategy also suits Markowitz type investors when we combine conservative strategies, such as Varspread, and more aggressive ones, such as Ensemble strategy, into a switching strategy.

For the purpose of tail loss protection, our proposed Ensemble strategy can be improved. Here, we measure risk with Value-At-Risk, nonetheless, other risk metrics can be considered for optimization, in particular Expected Shortfall or Maximum Drawdown. On top, it would be interesting to open the Ensemble learning tool box and to study how the different classifiers interacts which other in time, to better understand when their corresponding tail loss protection strategy might be suited for the respective market regime.

Acknowledgments We thank Nathalie Packham for her critical thoughts and support (Berlin School of Economics and Law, IRTG 1792 – Humboldt University Berlin).

References

- Black, F. 1976. Studies of stock price volatility changes. In *Proceedings of the 1976 meetings of the American Statistical Association*, 177–189. Business and Economics Statistics Section.
- Black, F., and R.W. Jones. 1987. Simplifying portfolio insurance. *J Portfolio Manag* 14 (1): 48–51.
- Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 (3): 307–327.
- Bollerslev, T., and V. Todorov. 2011. Tails, fears, and risk premia. *The Journal of Finance* 66 (6): 2165–2211.
- Bollerslev, T., V. Todorov, and L. Xu. 2015. Tail risk premia and return predictability. *Journal of Financial Economics* 118 (1): 113–134.
- Bouri, E., L. Gil-Alana, R. Gupta, and D. Roubaud. 2019. Modelling long memory volatility in the bitcoin market: Evidence of persistence and structural breaks. *International Journal of Finance & Economics* 24 (1): 412–426.
- Chen, S., C.Y.-H. Chen, W.K. Härdle, T. Lee, and B. Ong. 2018. Econometric analysis of a cryptocurrency index for portfolio investment. In *Handbook of blockchain, digital finance, and inclusion*, ed. D. Lee Kuo Chuen and R. Deng, vol. 1, 175–206. Academic Press.
- Christie, A. 1982. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics* 10 (4): 407–432.
- Christoffersen, P.F., and F.X. Diebold. 2006. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52 (8): 1273–1287.
- Chung, J., and Y. Hong. 2007. Model-free evaluation of directional predictability in foreign exchange markets. *Journal of Applied Econometrics* 22 (5): 855–889.
- Cizek, P., W.K. Härdle, and V. Spokoiny. 2009. Adaptive pointwise estimation in timeinhomogeneous conditional heteroscedasticity models. *The Econometrics Journal* 12 (2): 248–271.
- de Prado, M.L. 2018. *Advances in financial machine learning*. 1st ed. Wiley Publishing.
- Franke, J. 1999. Nonlinear and nonparametric methods for analyzing financial time series. In *Operations research proceedings 1998*, ed. P. Kall and H.-J. Lüthi, 271–282. Berlin Heidelberg: Springer.
- Franke, J., W.K. Härdle, and C.M. Hafner. 2019. *Statistics of financial markets – an introduction*. 5th ed. Springer Verlag.
- Hakansson, N. 1971. Capital growth and the mean-variance approach to portfolio selection. *Journal of Financial and Quantitative Analysis* 6 (1): 517–557.
- Hamidi, B., E. Jurzenko, and B. Maillet. 2009. A caviar modelling for a simple time-varying proportion portfolio insurance strategy. *Bankers, Markets and Investors* 102: 4–21.
- Happersberger, D., H. Lohre, and I. Nolte. 2019. Estimating portfolio risk for tail risk protection strategies. *European Financial Management*.
- Härdle, W.K., H. Herwatz, and V. Spokoiny. 2003. Time inhomogeneous multiple volatility modelling. *Journal of Financial Econometrics* 1: 55–99.
- Hillebrand, E. 2005. Neglecting parameter changes in garch models. *Journal of Econometrics* 129 (1): 121–138. Modelling structural breaks.
- Hochreiter, S., and J. Schmidhuber. 1997, November. Long short-term memory. *Neural Computation* 9 (8): 1735–1780.
- Hoerl, A.E., and R.W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12 (1): 55–67.
- Honegger, D., and D. Wijewickreme. 2013. Seismic risk assessment for oil and gas pipelines. In *Handbook of seismic risk analysis and management of civil infrastructure systems*, Woodhead Publishing Series in Civil and Structural Engineering, ed. S. Tesfamariam and K. Goda, 682–715. Woodhead Publishing.
- Jordà, O., and A.M. Taylor. 2011. *Performance evaluation of zero net-investment strategies*. Working paper 17150, National Bureau of Economic Research.
- Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (2): 263–291.

- Kelly, J.L. 1956. A new interpretation of information rate. *The Bell System Technical Journal* 35 (4): 917–926.
- Kim, H.Y., and C.H. Won. 2018. Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models. *Expert Systems with Applications* 103: 25–37.
- Kingma, D.P., and J. Ba (2015). *Adam: A method for stochastic optimization*, CoRR abs/1412.6980.
- Klochkov, Y., W. K. Härdle, and X. Xu (2019). Localizing multivariate CAViar. *SFB 649 Discussion Paper, submitted to Journal of Applied Econometrics* 2019 (007).
- Kozhan, R., A. Neuberger, and P. Schneider. 2013. The skew risk premium in the equity index market. *The Review of Financial Studies* 26 (9): 2174–2203.
- Krawczyk, B. 2016. Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence* 5 (4): 221–232.
- Kunreuther, H. 2002. Risk analysis and risk management in an uncertain world. *Risk Analysis* 22 (4): 655–664.
- Lambert, J., N. Matalas, C. Ling, Y. Haimes, and D. Li. 1994. Selection of probability distributions in characterizing risk of extreme events. *Risk Analysis* 14 (5): 731–742.
- Leshno, M., V.Y. Lin, A. Pinkus, and S. Schocken. 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks* 6 (6): 861–867.
- Linton, O., and Y.-J. Whang. 2007. The quantilogram: With an application to evaluating directional predictability. *Journal of Econometrics* 141 (1): 250–282.
- Longin, F., and B. Solnik. 2001. Extreme correlation of international equity markets. *The Journal of Finance* 56 (2): 649–676.
- Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7 (1): 77–91.
- Mason, S.J., J.S. Galpin, L. Goddard, N.E. Graham, and B. Rajaratnam. 2007. Conditional exceedance probabilities. *Monthly Weather Review* 135 (2): 363–372.
- McNeil, A., and R. Frey. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7 (3): 271–300. Special issue on Risk Management.
- Mercurio, D., and V. Spokoiny. 2004. Statistical inference for time-inhomogeneous volatility models. *The Annals of Statistics* 32: 577–602.
- Packham, N., J. Papenbrock, P. Schwendner, and F. Woebbecking. 2017. Tail-risk protection trading strategies. *Quantitative Finance* 17 (5): 729–744.
- Rickenberg, L. 2019. Tail risk targeting: Target var and cvar strategies. *SSRN Electronic Journal*.
- Spokoiny, V. 1998. Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics* 26: 1356–1378.
- . 2009. Multiscale local change point detection with applications to value-at-risk. *The Annals of Statistics* 37: 1405–1436.
- Spokoiny, V., and M. Zhilova. 2015. Bootstrap confidence sets under model misspecification. *The Annals of Statistics* 43 (1): 2653–2675.
- Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15 (56): 1929–1958.
- Taylor, J., and K. Yu. 2016. Using auto-regressive logit models to forecast the exceedance probability for financial risk management. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179 (4): 1069–1092.
- Wen, F., Z. He, X. Chen, and J. Li. 2014. Investors' risk preference characteristics and conditional skewness. *Mathematical Problems in Engineering*.
- Zhang, Z., S. Zohren, and S. Roberts. 2020. Deep learning for portfolio optimization. *The Journal of Financial Data Science*.



Yiannis Karavias

Abstract

The purpose of this chapter is to review recently developed methods for analyzing structural breaks in panel data. Structural breaks are caused by events that change the parameters of economic and financial models, such as the 2007–2008 financial crisis and the COVID-19 crisis. If not correctly accounted for they lead to an invalid inference. It is not always clear if a significant event has affected a particular economic model or relationship, or the date on which the relationship suffered the change. Answers to these questions can only be given through formal econometric analysis. This chapter reviews methods for both detecting and dating structural breaks. Furthermore, the usefulness of structural break method is demonstrated in examining trade credit determinants in a panel of small and medium European enterprises. A structural break was detected and dated in the year 2010. Before the break, firm profitability, sales growth, and bank loans all contributed to an increase in trade credit. After the break, the bank loans' effect was significantly reduced, providing evidence that the trade credit's redistribution channel collapsed after the 2007–2008 financial crisis.

Keywords

Panel data · Structural change · Cross-section dependence · Structural break · Trade credit · Financial crisis

JEL Classification

C12 · C13 · C33

Y. Karavias (✉)

Department of Economics, University of Birmingham, Birmingham, UK

e-mail: i.karavias@bham.ac.uk

93.1 Introduction

Structural changes or “breaks” are recurring phenomena in economic and financial systems. They are unanticipated shocks that can cause significant changes in the established relationships between key variables. This has the potential to cause expectations and forecasts formed from econometric models to go significantly astray. For example, investment decisions made before the 2007–2008 financial crisis or the 2020 COVID-19 crisis were based on very different expectations and forecasts to those obtained after these crises.

While structural changes are sometimes clearly observed, at other times, they can occur more subtly and can only be detected later by formal econometric methods. Their recurring nature makes their study and identification a timely issue. Current practice detects breaks by analyzing time series data. Still, detection is much more precise when it is done on information-rich panel data. A review of structural break applications in financial time series can be found in Andreou and Ghysels (2009).

Panel data are complex data sets, which are being used to study interconnected assets, firms, and markets at a local, national, and international level. They provide a unique capacity for modeling the mosaic of heterogeneous and interlinked economies, as they contain information not available in a single time series. Some examples of their application in financial economics include bank profitability, see, for example, Delis and Karavias (2015), the determinants of board structure, see, for example, Wintoki et al. (2012), multi-factor CAPM, see, for example, Makwasha et al. (2019), firm performance, see, for example, Asimakopoulous et al. (2020), firm employment, see, for example, Da Silva Fernandes et al. (2019), forecasting, see, for example, Liu et al. (2020), and return predictability see, for example, Westerlund et al. (2017).

The objectives of structural change analysis are to find answers to questions such as “Was there a structural change?” and if so, “When did the change happen?”. Answering these questions may help identify those events which have caused the structural change and inform future expectations and investment decisions. Where a break is detected, [estimating the date of the break](#) facilitates tracing the cause of change, making similar structural breaks easier to predict in the future. Connecting a structural break to an event can happen by matching the break’s estimated date to that event. Furthermore, knowledge of the existence of two distinct sample periods, pre-break and post-break, will allow the correct measurement of economic outcomes in each period. Finally, when models are subject to structural change, better forecasts can be obtained by putting increased weight on observations after the break, see, for example, Pesaran and Timmermann (2004).

The first aim of this chapter is to review methods for the analysis of structural breaks in panel data. Such techniques were first developed about a decade ago. Bai (2010) was the first paper to derive an estimator for the date of the break in a panel data model with individual intercepts but no other regressors. The first test for structural change in panel data was developed by De Wachter and Tzavalis (2012),

for dynamic panel data models with exogenous or predetermined regressors. Ever since, the literature has gained traction with many new methods being proposed, see, for example, Kim (2014), Baltagi et al. (2016), Hidalgo and Schafgans (2017), and Karavias et al. (2021). Structural breaks have also been considered in panel data unit root and cointegration tests, see, for example, Westerlund (2006), Bai and Carrion-i-Silvestre (2009), Hadri et al. (2012), Banerjee and Carrion-i-Silvestre (2015), and Karavias and Tzavalis (2014, 2017, 2019). The literature review in this chapter will focus on methods for detecting and dating structural breaks and not on panel data unit root and cointegration tests with breaks, although some of these methods will be used in the application part.

The second aim of the chapter is to demonstrate the usefulness of the structural break analysis. This is done by examining trade credit provision. Meltzer (1960) was the first paper to examine the determinants of trade credit. Ever since, this literature has flourished, with the most recent contributions being Box et al. (2018) and Coricelli and Frigerio (2019). Trade credit is the most important source of short-term firm finance with implications at both the microeconomic level, see, for example, Asimakopoulos et al. (2020) on the impact of trade credit on firm performance, or at the macro level see, for example, Meltzer (1960) for the impact of trade credit on monetary policy and Fisman and Love (2003) for the impact of trade credit on economic growth.

Using a large sample of Small and Medium Enterprises (SME's) over the period 2004–2015, this chapter examines how trade credit is determined by sales growth, firm profits, and bank loans. The initial regression, which does not account for structural breaks, strikingly finds that bank loans are insignificant to the amount of trade credit provided, and therefore, that there is no redistribution of bank funds to firms that have no direct access to bank lending. Testing for structural change, however, reveals that a break existed in 2010. The detection of the break and the estimation of its date is done using the methodology of Karavias et al. (2021). The latter is the first paper to establish the complete asymptotic theory for testing and dating structural breaks in panel data.

Once this break is incorporated in the trade credit regression, the new results show that bank lending had a positive and statistically significant effect on trade credit before 2010, but this effect becomes zero after the break. This finding provides evidence of the redistribution channel of trade credit before 2010; see, for example, Fisman and Love (2003). It further shows that the Great Recession reached European SME's in 2010 and stopped the redistribution of bank lending. Finally, the relationship between firm profits and trade credit remains constant in time and is unaffected by structural breaks. The same applies to trade credit generated by increased sales.

The chapter is organized as follows. Section 93.2 introduces structural breaks in panel data models. Section 93.3 reviews the literature on hypothesis tests for the existence of structural change. Section 93.4 reviews the literature on dating structural breaks. Section 93.5 analyses structural breaks in the provision of trade credit. Finally, Sect. 93.6 concludes.

93.2 Structural Breaks in Panel Data: The Model

Consider the general panel data model with N units and T periods, that is $i = 1, \dots, N$, and $t = 1, \dots, T$, where a break has occurred at time k :

$$y_{i,t} = \begin{cases} \beta_i' x_{i,t} + e_{i,t}, & \text{for } t \in \{1, \dots, k\} \\ \beta_i' x_{i,t} + \delta_i' z_{i,t} + e_{i,t}, & \text{for } t \in \{k+1, \dots, T\}, \end{cases}$$

The above model may be written more compactly as:

$$y_{i,t} = \beta_i' x_{i,t} + \delta_i' z_{i,t}(k) + e_{i,t}, \quad (1)$$

where $x_{i,t}$ is a $p \times 1$ vector of regressors, $z_{i,t}(k)$ is a $q \times 1$ vector, $e_{i,t}$ is the error term and $y_{i,t}$ is the scalar dependent variable. The coefficients β_i and δ_i are $p \times 1$ and $q \times 1$ vectors, respectively. The regressor $z_{i,t}(k)$ is defined as:

$$z_{i,t}(k) = R' x_{i,t} I(t > k), \quad (2)$$

where R is a $p \times q$ matrix with full column rank q , $I(\cdot)$ is the indicator function, and k is the date of the break. The matrix R selects the subset of q regressors in $x_{i,t}$ which experience a change in coefficient at time k . If $p > q$, then $R = (0_r \times (p - q), I_q)'$, while if $p = q$ then $R = I_p$ and all the coefficients change at the time of the break. I_q and I_p are q and p dimensional identity matrices, respectively. For $p = q$ the model is called a “pure structural change model,” while for $q < p$ the model is called a “partial structural change model.”

The aim in the analysis of structural change is two-fold: (a) test for the existence of a structural break, and (b) estimate the date of the break and construct a confidence interval which contains the true break date with probability $(1 - \alpha)\%$, where α is the significance level.

The above model allows for a common break which happens at the same time k for all units. The common break assumption is acceptable in most settings and largely depends on the frequency of the data. It is generally easier to justify in low-frequency data. For example, changes in economic policy will appear in a specific year in annual balance sheet data, even though each firm may have responded on a different day or month to the change in policy. A common break can also be justified as the mean of random breaks, see, for example, Bai (2010).

Testing for structural change is equivalent to testing the null hypothesis that $H_0 : \delta_i = 0$, for all i , against one of two different alternative hypotheses. The first alternative hypothesis is for the scenario where the researcher wishes to test the null against the alternative hypothesis that the break occurred at a selected date. This alternative can be written as $H_1(k_0) : \delta_i \neq 0$ for some $i \in \{1, \dots, N\}$. In this case, the true break date k_0 is assumed known by the researcher. If the investigator is mistaken on this choice, the resulting structural change test will not behave properly; specifically, it will not have size $\alpha\%$. Consider Brexit, for example. Did its effect materialize a day after the 2016 referendum, possibly through changes in economic

agents' expectations? Did the effect emerge when the UK left the EU? Will it materialize after the transition period is over? This example shows that even if the researcher has an idea of the break's date, the date may not be correct. For these reasons, we would like to have a way of doing the above hypothesis test without assuming a known break date. The second and more empirically relevant hypothesis is that the break may have occurred at an unknown date. This alternative can be stated as $\cup_{k \in K} H_1(k) : \delta_i \neq 0$ for some $i \in \{1, \dots, N\}$, where K is the set of all possible break dates.

If the date of the break is unknown and the test rejects the null hypothesis, then it becomes of interest to estimate the date of the break k_0 and construct a confidence interval which contains the true break date with probability, say 95%, depending on our significance level. For this, we need an estimator $\hat{k}$ and its limiting distribution. Unfortunately, inference in both cases is non-standard as neither the test for structural change against the alternative of unknown break date or the break date estimator have the typical limiting distributions, for example, standard normal or χ^2 , and therefore these must be either tabulated if they are free of nuisance parameters or bootstrapped if they are not. The following sections demonstrate how this can happen.

93.3 Structural Breaks in Panel Data: The Errors

The specification of the error term completes the model (1), where $e_{i,t}$, which is defined as:

$$e_{i,t} = \gamma_i' f_t + \varepsilon_{i,t}, \quad (3)$$

where γ_i is an $m \times 1$ vector of individual-specific factor loadings and f_t is an $m \times 1$ vector of common factors. If $m = 1$ and $f_t = 1$ for all t and the model is homogeneous, that is, $\beta_i = \beta$ and $\delta_i = \delta$ for all i , we have the typical panel data model with individual effects which may be fixed or random. The γ_i This case is called unobserved individual heterogeneity and captures systematic differences in the units, which do not change over time. For example, if the units are countries, individual heterogeneity captures the effects of geography, culture, and state efficiency. If the units are firms, then heterogeneity captures technological differences, years of education, and managers' ability, among others. The unobserved heterogeneity is usually correlated with the regressors because the regressors depend on the characteristics mentioned above, see, for example, Baltagi (2008) for further details.

If $m = 2$, $\gamma_i = (\xi_i, 1)'$ and $f_t = (1, \psi_t)'$ where ξ_i and ψ_t are scalars, then $\gamma_i' f_t = \xi_i + \psi_t$, which is the two-way error component model with individual heterogeneity and time effects, see, for example, Baltagi (2008). Time effects capture aggregate time-series effects such as the oil price, economic growth, and financial volatility. These are common shocks at time t , which affect all units simultaneously. The error model with $m = 2$ is the simplest model of cross-section dependence between the units in the panel. Cross-section dependence is a critical issue in financial and economic data,

see, for example, Sarafidis and Wansbeek (2012). If $\gamma_i' f_t = \xi_i + \psi_t$ then, consistent parameter estimation requires that these effects are removed before estimation. The removal can be done easily due to the simple structure of the error components by subtracting the cross-sectional averages for ψ_t and the within-groups mean for ξ_i .

More general panel specifications consider the “interactive effects” model with m factors: $\gamma_i' f_t$. In this model, the γ_i loadings capture the heterogeneous responses of individual units to the common shocks captured by the factors f_t . The interactive effects panel data models are widely applicable. Consider, for example, the Fama and French (1992) model. The dependent variable $y_{i,t}$ is the excess return of asset i at period t , and $x_{i,t}$ is a set of observable factors such as dividend yields or balance sheet data on the five factors of Fama and French (2015). The factors f_t in the error term are unobservable factors left out of the model specification, see, for example, Cochrane (2011). Another example comes from the firm performance literature, where $y_{i,t}$ is a measure of firm performance such as the return on assets, $x_{i,t}$ captures observable firm characteristics such as liquidity, debt, and sales, and the unobserved factors f_t capture the effects of changes in technology, which are heterogeneous for each firm according to the factor loadings γ_i .

It is reasonable to assume that these factors are correlated with the regressors through the equation:

$$x_{i,t} = \Gamma_i' f_t + v_{i,t}, \quad (4)$$

where Γ_i is an $m \times p$ matrix of factor loadings and $v_{i,t}$ is a $p \times 1$ vector of idiosyncratic errors. The correlation between $x_{i,t}$ with $e_{i,t}$ through the interactive effects $\gamma_i' f_t$ makes $x_{i,t}$ endogenous, consistent estimators must be used. The two more popular are the concentrated least squares estimator of Bai (2009) and the Common-Correlated Effects (CCE) estimator of Pesaran (2006).

93.4 Testing for Structural Change

Testing for structural change when the date of the break is known, that is, testing H_0 against $H_1(k_0)$ can be done easily through a Wald or F test as in Chow (1960). In the following, we shall denote such a test statistic as $W(k_0)$. The limiting distribution of $W(k_0)$ is a χ_q^2 distribution with q degrees of freedom, for which critical values are readily available.

Testing H_0 against $\cup_{k \in K} H_1(k)$ can not be done using the typical Wald tests because the break's date is a parameter that appears only under the alternative. This problem was studied by Davies (1977, 1987) who proposed using the following test statistic:

$$SW = \sup_{k \in K} W(k),$$

derived from the union-intersection principle, see, for example, Casella and Berger (2001, p. 380). The idea behind this statistic is that if $W(k)$, for each $k \in K$, is greater

than an appropriate critical value, then we can reject the null hypothesis. In other words, if the supremum of the $W(k)$ is greater than a critical value, the test is rejected. Andrews (1993) derived the asymptotic distribution of SW for single time series models and tabulated its critical values. The researcher determines the set of permissible break dates K . Some trimming is necessary for the first and last time periods, so that there are enough data to identify the pre- and post-break regimes. Andrews (1993) suggests that $K = 0.15T, 0.85T$, where $\lfloor \cdot \rfloor$ is the greatest integer function.

An alternative method for deriving structural change tests is using the cumulative sum of squares test (CUSUM) in Brown et al. (1975). However, this test, and its variants, have several pronounced weaknesses. They suffer from trivial power against some types of alternatives, see, for example, Kramer et al. (1988), they have nonmonotonic power functions, see, for example, Andreou and Ghysels (2009), and ultimately, they have worse power and size performance when compared to SW , see, for example, Andrews (1993).

De Wachter and Tzavalis (2012) were the first to consider testing structural changes in panel data. They maintain a homogeneous panel model with lagged dependent and exogenous variables as regressors and individual fixed effects. Their $W(k)$ statistic is based on $\hat{\beta}$ and $\hat{\delta}$ estimators, which is estimated by the generalized method of moments as in Arellano and Bond (1991). Zhu et al. (2020) extend this test to allow for errors as in (3). Both these tests have excellent size and power properties when $N \rightarrow \infty$ and T are fixed. However, the limiting distribution must be simulated every time, which hinders their application. Another issue arises when T is moderate, in which case the small sample behavior deteriorates as they have to estimate many parameters. CUSUM tests are considered by Antoch et al. (2019) and Hidalgo and Schafgans (2017), under more restrictive error assumptions. Antoch et al. (2019) assume that the interactive effects are asymptotically negligible, while Hidalgo and Schafgans (2017) considered a weaker form of cross-section dependence than the one which appears in (3). Both tests need simulated critical values for each application. Karavias et al. (2021) consider testing for breaks in the model (1), (2), (3), and (4) with $\beta_i = \beta$ and $\delta_i = \delta$ using the SW statistic based on a new variant of the pooled CCE estimator of Pesaran (2006). They show that the limiting distribution of SW , in that case, coincides with that of Andrews (1993). Therefore, the same critical values can be used for hypothesis testing. Additionally, it was found that the test has excellent size and power properties, even for small T . This test will be used in Sect. 93.5, in the empirical part of this chapter.

93.5 Estimating the Date of the Break

Given the existence of a structural break, the next step is to estimate the break's date and construct a valid confidence interval. The estimator of choice in the literature is:

$$\hat{k} = \underset{k \in K^*}{\operatorname{argmin}} SSR(k),$$

which is the breakpoint which minimizes the sum of squared residuals. The method of calculating this estimator is as follows. Step 1: Select the first possible break date

$k \in K^*$, say k^1 . Then, considering k^1 given, estimate the model and get the $SSR(k^1)$. Step 2: repeat this process for all $k \in K^*$. Step 3: find the minimum of all $SSR(k)$; then $\hat{k}$ is the estimate which corresponds to this minimum. Notice that $K^* = [q, T - q - 1]$, is different from K .

The break date estimator exhibits an irregular behavior regarding its rate of convergence and its asymptotic distribution. This behavior has an impact on its asymptotic distribution. Bai (2010) considered the heterogeneous location model where $x_{i,t} = 1$, with independent errors, and derived the break date estimator's consistency and asymptotic distribution. Kim (2011, 2014) considered the model with intercept and linear trend $x_{i,t} = (1, t)'$ but also errors as in (3). Similarly, he derived consistency, rate of convergence, and asymptotic distribution. Kim (2011, 2014) shows that the convergence rate depends on the interplay between the error structure and the factor removal method.

Baltagi et al. (2017) consider a homogeneous version of (1) with independent errors that may be stationary or not and exogenous regressors that may be stationary or not. They prove the consistency and derive the rate of convergence and the asymptotic distribution of $\hat{k}$. The main finding is that if $N \rightarrow \infty$ and $T/N \rightarrow \infty$ then $\hat{k}$ is consistent regardless of the stationarity properties of $e_{i,t}$ and $x_{i,t}$. When both the regressor and the errors are nonstationary, then the rate of convergence is very fast, of the order $O((NT)^{-1})$. Further consistency results are provided by Boldea et al. (2020) which for the first time consider the fixed T case in a homogeneous model with independent errors. They show that $\hat{k}$ based on residuals of the ordinary least squares estimator, the estimator, which does not remove the individual effects in the error term, is consistent.

For the full model (1), (2), (3), and (4) with exogenous regressors and common factors in the errors, Baltagi et al. (2016) prove the consistency of $\hat{k}$ when the model is estimated by the CCE estimator of Pesaran (2006). Baltagi et al. (2019) extend this consistency result to endogenous regressors. Li et al. (2016) consider the homogeneous version of the above, where $\beta_i = \beta$ and $\delta_i = \delta$ for all i , and prove the consistency of $\hat{k}$, but use principal components to estimate the common factors, instead of CCE. Karavias et al. (2021) also consider the homogeneous version of (1), and they prove the consistency of $\hat{k}$ and derive its rate of convergence and its limiting distribution. They find that the convergence rate depends on the relationship between the number of regressors p and factors m . If $m < p$ and N is larger than T , then the speed of convergence to the true k_0 increases. Furthermore, they show that consistency and the limiting distribution hold when both T is fixed or asymptotic, and derive valid confidence intervals for $\hat{k}$ in both cases. All these papers consider common breaks for all units. Okui and Wang (2020) have considered estimation in panel data where breaks are common within a group, and there are several groups of units in the panel.

93.6 Structural Breaks in Trade Credit Provision

As it was argued in the introduction, financial relationships are subject to structural breaks frequently enough, so that it is suggested that testing for breaks should be reported together with descriptive statistics in any research paper. If a break is

detected, this will impact the rest of the econometric analysis. Karavias et al. (2021), that is, examine the impact of COVID-19 on stock returns using weekly data on sixty-one countries. They find that COVID-19-related variables such as cases, deaths, and government restrictions hurt stock returns during the first three months of the pandemic. A structural break happens, and the adverse effects disappear; the stock market is no longer affected by the COVID-19-related variables. The break is estimated to take place in the first week of April 2020. The result is attributed to the massive quantitative easing programs put forth by central banks across the world in the last days of March 2020.

In this chapter, we wish to examine if there is a structural break in the fundamental relationship between trade credit provision and its determinants, particularly a change related to the 2007–2008 financial crisis. Trade credit is a type of short-term loan that supplier firms provide to customer firms at the same time as goods. Suppose customers cannot pay for the goods upon delivery. In that case, they enter a contract with the supplier to delay payment for a while, with interest. Such agreements are usually standardized. An example is a contract named “2-10 net 30,” which means that if the customer pays in the next ten days after the delivery of goods, they will get a 2% discount. Still, after the first ten days, they lose the discount and have to pay the full amount within 30 days of delivery. The 2% discount can be seen as a loan with an interest rate of 2% for 20 days. If this is annualized, then the one-year interest rate is 44%. Because trade credit is so expensive, many theories are trying to explain why it exists. Petersen and Rajan (1997) provide an excellent review of the topic.

Two theories which are of particular interest are the financial motivation theory of Emery (1984) and the redistribution view of trade credit of Meltzer (1960). The financial motivation theory posits that suppliers have more information and power over customer firms and, therefore, can offer credit at better rates than banking institutions. Customer firms prefer this expensive credit over the even more costly credit of financial institutions. The redistribution view of trade credit argues that firms with access to cheap bank finance will use this financial motive to redistribute bank credit to the less privileged firms.

The theories mentioned in the previous paragraph can be tested using the following panel data regression model:

$$TC_{i,t} = \beta_1 S_{i,t} + \beta_2 L_{i,t} + \beta_3 R_{i,t} + \beta_4 B_{i,t} + \beta_5 SG_{i,t} + e_{i,t}, \quad (5)$$

where $TC_{i,t}$ is the trade credit extended by firm i at time t , $S_{i,t}$ is the size of the firm, $L_{i,t}$ is firm liquidity, $R_{i,t}$ are firm profits, $B_{i,t}$ is bank loans received by the firm, and $SG_{i,t}$ is sales growth since the previous period. This is a standard regression in the literature, see, for example, Meltzer (1960), Nadiri (1969), Petersen and Rajan (1997), and Box et al. (2018). The parameters of interest are the coefficients of profits, sales growth, and bank loans. Positive profit and sales growth coefficients mean that an increase in these variables generates trade credit directly; therefore, suppliers provide trade credit because they have a financial motive. On the other hand, a positive coefficient of bank loans means that suppliers redistribute bank loans to their customers in trade credit.

It is assumed that the errors $e_{i,t}$ include common factors as in (3), capturing firm characteristics, industry, and country effects, as well as the impact of other unobserved factors such as consumer confidence, interest rate conditions, economic growth, and oil price fluctuations.

The data includes observations on 2526 European SMEs over the period 2004–2015. The data came originally from Amadeus (Analyse Major Database from European Sources) database, distributed by Bureau Van Dijk (BvDEP). A description of the data collection and cleaning can be found in Asimakopoulou et al. (2020). The variable $TC_{i,t}$ is equal to the accounts receivables over total assets. Size $S_{i,t}$ is set equal to the logarithm of total firm assets, $L_{i,t}$ is defined as cash and cash equivalents over total assets, $R_{i,t}$ is equal to the returns on assets before taxes, $B_{i,t}$ is defined as bank loans over total assets, and finally, $SG_{i,t}$ is sales growth of the firm from period $t - 1$ to period t .

Table 1 provides descriptive statistics for the variables above. The average trade credit in the sample is 33% of the total assets, signaling the item's significance in the firm balance sheets. Figure 1 presents the trade credit provided by the average firm by year. A significant drop can be seen for the year 2009. Afterward, it seems that trade credit fluctuates along this new level. Given that the data include a small set of time series observations, two fixed-T panel unit root tests are employed, the Harris and Tzavalis (1999) test and the Karavias and Tzavalis (2019) test. The results appear in the HT and KT19 columns. The specifications allow for intercepts only and KT19 allows for a structural break in the series' mean under the alternative. The results are in favor of stationarity for all series.

The first column of Table 2 presents the regression results (5), where no structural breaks have been considered. The estimation method used is that of Pesaran (2006), with no observed common factors. Furthermore, the matrix of cross-section averages contains only the averages for the variables $S_{i,t}$, $L_{i,t}$, $R_{i,t}$, $B_{i,t}$, and $SG_{i,t}$, and not the averages of $TC_{i,t}$ as in Pesaran (2006), as in Karavias et al. (2021). Profits, bank loans, and sales growth all have the expected positive signs. From the three, bank loans are statistically insignificant, meaning no loan redistribution is taking place. The other two are statistically significant. Size is negative but insignificant. Liquidity has a negative and statistically significant effect, a result in line with the literature, see, for example, Meltzer (1960), and Coricelli and Frigerio (2019).

It is time now to examine the above relationship for structural breaks. A modeling issue which arises is how to choose which variables are subject to a structural break. There is no compelling argument on why size and firm liquidity should have different effects based on the potential impact of the financial crisis, so they are assumed unchanged. This is also corroborated by unreported robustness checks. Concerning the remaining three variables, for robustness, the following three regressions are examined:

$$TC_{i,t} = \beta_1 S_{i,t} + \beta_2 L_{i,t} + \beta_3 R_{i,t} + \beta_4 B_{i,t} + \beta_5 SG_{i,t} + \delta_1 SG_{i,t} I(t > k) + e_{i,t}, \quad (6)$$

Table 1 Descriptive statistics

Variable	Obs.	Mean	SD	Min	Max	HT		KT19		Break
						AR(1)	P-value	AR(1)	P-value	
Trade credit	30,312	0.334	0.153	0	0.909	0.522	0	0.954	0	4
Size	30,312	8.167	1.141	5.336	11.35	0.753	0	1.004	1	
Liquidity	30,312	0.149	0.131	0	0.635	0.487	0	0.939	0	8
Profitability	30,312	0.075	0.075	-0.229	0.369	0.364	0	0.803	0	8
Bank loans	30,312	0.07	0.085	0	0.578	0.386	0	0.862	0	8
Sales growth	30,312	0.035	0.179	-2.667	3.109	-0.178	0	-0.057	0	8

Notes: This table reports selected descriptive statistics for a panel of 2526 firms observed over 2004–2015. For each variable, the mean, standard deviation (SD), minimum (Min), and maximum (Max) statistics are reported. The HT test refers to the Harris and Tzavalis (1999) homogeneous-coefficient panel unit root test. The reported first-order autoregression (AR(1)) coefficient is bias-adjusted, and its associated p-value is also reported. The KT19 test refers to the Karavias and Tzavalis (2019) fixed-T test that allows for serial correlation and a common break under the alternative; we assume uncorrelated errors

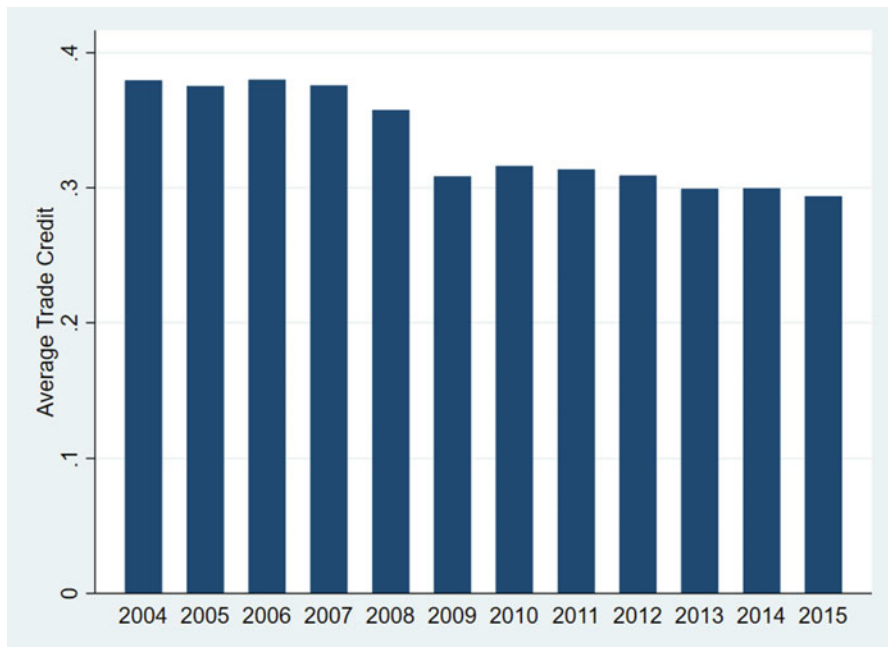


Fig. 1 Average trade credit over time

$$TC_{i,t} = \beta_1 S_{i,t} + \beta_2 L_{i,t} + \beta_3 R_{i,t} + \beta_4 B_{i,t} + \beta_5 SG_{i,t} + \delta_1 SG_{i,t}I(t > k) + \delta_2 B_{i,t}I(t > k) + e_{i,t}, \quad (7)$$

$$TC_{i,t} = \beta_1 S_{i,t} + \beta_2 L_{i,t} + \beta_3 R_{i,t} + \beta_4 B_{i,t} + \beta_5 SG_{i,t} + \delta_1 SG_{i,t}I(t > k) + \delta_2 B_{i,t}I(t > k) + \delta_3 R_{i,t}I(t > k) + e_{i,t}, \quad (8)$$

which correspond to model (1) with $p = 5$ and $q = 1, 2$ and 3 , respectively.

Columns (2), (3), and (4) of Table 2 present the above regression results. Column (2) considers regression (6). The *SW* test of Karavias et al. (2021) rejects the null hypothesis of no structural change at the 10% significance level. The $\hat{k}$ estimator of the same paper estimates the date of the break in 2010 with a 95% confidence interval being (2009, 2011). This seems to be reasonable as there was a delay in the propagation of the global financial crisis in Europe. The coefficient of sales growth before the crisis is positive and statistically significant, but after the crisis, it is reduced by δ_1 . The coefficient after the crisis is equal to $\hat{\beta}_5 + \hat{\delta}_1 = 0.055 - 0.01 = 0.045$, which is still positive. Notice that δ_1 is barely not significant at the 10% level. This result supports the financial motive for firms.

Column (3) considers the same regression, but now the bank loans variable changes coefficient as well. The *SW* test now strongly rejects the null of no structural

Table 2 Main regression results

Regression	(1)	(2)	(3)	(4)
Size	−0.009 (0.005)	−0.01 (0.005)**	−0.001 (0.005)	−0.005 (0.005)
Liquidity	−0.454 (0.012)***	−0.45 (0.009)***	−0.475 (0.010)***	−0.473 (0.010)***
Profitability − β_3	0.204 (0.012)***	0.204 (0.010)***	0.2 (0.011)***	0.205 (0.015)***
Profitability − δ_3				−0.009 (0.023)
Bank loans − β_4	0.012 (0.014)	0.013 (0.011)	0.049 (0.015)***	0.058 (0.015)***
Bank loans − δ_2			−0.047 (0.022)**	−0.069 (0.025)***
Sales growth − β_1	0.048 (0.004)***	0.055 (0.005)***	0.054 (0.005)***	0.054 (0.005)***
Sales growth − δ_1		−0.01 (0.007)	−0.001 (0.007)	−0.012 (0.008)
<i>SW</i>		7.412*	17.978***	32.281***
Break date		2010	2010	2010
95% CI		[2009, 2011]	[2009, 2011]	[2009, 2011]
Obs.	30,312	30,312	30,312	30,312
Firms	2526	2526	2526	2526
Years	12	12	12	12
Interactive effects	Yes	Yes	Yes	Yes

Notes: This table is based on regressions containing a panel of 2526 firms, where the dependent variable is $TC_{i,t}$, and control variables are $S_{i,t}$, $L_{i,t}$, $R_{i,t}$, $B_{i,t}$, and $SG_{i,t}$. *SW* is the Sup-Wald test for the existence of a structural break. Finally, *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively

breaks, and the break date is the same as before. Before the break, bank loans had a positive and statistically significant effect on trade credit at the 1% level. This is the first evidence in favor of the redistribution channel of trade credit. After the break, the bank loan coefficient becomes $\hat{\beta}_4 + \hat{\delta}_2 = 0.049 - 0.047 = 0.02$, which is very close to zero. Notice that $\hat{\delta}_2$ is statistically different from zero and the test for $H_0 : \beta_4 = -\delta_2$ has a p-value equal to 1. All this is evidence that after 2010, the redistribution of bank loans stopped. The conclusion here is that, if structural breaks were not considered, there would be no evidence of the redistribution channel of trade credit. Still, the structural breaks analysis shows that this channel is potent up until 2010. The coefficients of sales growth are similar to column (3). Column (4) of the table confirms these results and also shows that the break does not impact the effect of profitability on trade credit because, although the coefficient $\hat{\delta}_3$ is negative, it is small and insignificant. Overall, the structural break analysis allows us to conclude that the financial motivation theory holds and the redistribution theory of trade credit holds until 2010 but not later.

93.7 Conclusions

This chapter's first contribution is a review of the panel data structural breaks detection and estimation literature. This literature has considered a diverse set of panel data models and has found significant, qualitatively new results. Despite the tremendous progress achieved in the past decade, there are still many questions open for research in nonstationary panels, panels with group structure, and models with multiple structural breaks.

The second contribution of the paper is the analysis of the determinants of trade credit. Meltzer (1960) argues that banks reduce lending in times of contractionary monetary policy; therefore, bank loans are available only to the largest and more trustworthy firms. In turn, these firms will redistribute the available credit in the form of trade credit to their customers. Meltzer (1960) argues that this process reduces the adverse impact of money tightening on small firms. In this chapter, using data on a large sample of European firms for the period 2004–2015, we find evidence that this theory is valid up to a point. A structural break was detected and dated in 2010. Before 2010, the relationship between bank loans and trade credit was positive, which means that some of the bank funds were being redistributed in the form of trade credit. However, after 2010, this phenomenon stopped.

To summarize, structural change has long been recognized as one of the greatest threats to inference. The longer the time dimension of the panel, the higher the probability that a break has happened. A suggestion would be that tests for structural breaks are routinely reported with other descriptive statistics in papers with panel data regressions.

93.8 Cross-References

► [Time-Series and Cross-Sectional Tests of Asset Pricing Models](#)

References

- Andreou, E., and E. Ghysels. 2009. Structural breaks in financial time series. In *Handbook of financial time series*, ed. T. Mikosch, J.P. Kreib, R. Davis, and T. Andersen. Berlin/Heidelberg: Springer.
- Andrews, D.W.K. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61: 821–856.
- Antoch, J., J. Hanousek, L. Horvath, M. Huskova, and S. Wang. 2019. Structural breaks in panel data: Large number of panels and short length time series. *Econometric Reviews* 38: 828–855.
- Arellano, M., and Bond, S. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58 (2): 277–297.
- Asimakopoulos S., F. Da Silva Fernandes, and Y. Karavias. 2020. *Firm heterogeneity and trade credit behaviour*. Discussion papers 20-20, Department of Economics, University of Birmingham.
- Bai, J. 2009. Panel data models with interactive fixed effects. *Econometrica* 77: 1229–1279.

- . 2010. Common breaks in means and variances for panel data. *Journal of Econometrics* 157: 78–92.
- Bai, J., and J.L. Carrion-I-Silvestre. 2009. Structural changes, common stochastic trends, and unit roots in panel data. *Review of Economic Studies* 76: 471–501.
- Baltagi, B.H. 2008. *Econometric analysis of panel data*. Chichester: Wiley.
- Baltagi, B.H., Q. Feng, and C. Kao. 2016. Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics* 191: 176–195.
- Baltagi, B.H., C. Kao, and L. Liu. 2017. Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term. *Econometric Reviews* 36: 85–102.
- Baltagi, B.H., Q. Feng, and C. Kao. 2019. Structural changes in heterogeneous panels with endogenous regressors. *Journal of Applied Econometrics* 34: 883–892.
- Banerjee, A., and J.L. Carrion-i-Silvestre. 2015. Cointegration in panel data with structural breaks and cross-section dependence. *Journal of Applied Econometrics* 30 (1): 1–23.
- Boledea, O., B. Drepper, and Z. Gan. 2020. Change point estimation in panel data with time-varying individual effects. *Journal of Applied Econometrics* 35: 712–727.
- Box, T., R. Davis, M. Hill, and C. Lawrey. 2018. Operating performance and aggressive trade credit policies. *Journal of Banking & Finance* 89: 192–208.
- Brown, R., J. Durbin, and J. Evans. 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B (Methodological)* 37 (2): 149–192.
- Casella, G., and Berger, R. 2001. *Statistical Inference*. 2nd Edition, Duxbury Press, Pacific Grove, 363.
- Chow, G.C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28 (3): 591–605.
- Cochrane, John H. 2011. Presidential address: Discount rates. *Journal of Finance* 66 (4): 1047–1108.
- Coricelli, F., and M. Frigerio. 2019. Interenterprise credit and adjustment during financial crises: The role of firm size. *Journal of Money, Credit and Banking* 51: 1547–1580.
- Davies, R.B. 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64: 247–254.
- . 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74: 33–43.
- De Watcher, S., and E. Tzavalis. 2012. Detection of structural breaks in linear dynamic panel data models. *Computational Statistics and Data Analysis* 56: 3020–3034.
- Delis, M.D., and Y. Karavias. 2015. Optimal versus realized bank credit risk and monetary policy. *Journal of Financial Stability* 16 (C): 13–30.
- Emery, G.W. 1984. A pure financial explanation for trade credit. *Journal of Financial and Quantitative Analysis* 19: 271–285.
- Fama, E.F., and K.R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47 (2): 427–465.
- . 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116: 1–22.
- Fernandes, F.D.S., A. Kontonikas, and S. Tsoukas. 2019. On the real effect of financial pressure: Evidence from firm-level employment during the euro-area crisis. *Oxford Bulletin of Economics and Statistics* 81 (3): 617–646.
- Fisman, R., and I. Love. 2003. Trade credit, financial intermediary development and industry growth. *Journal of Finance* 58: 353–374.
- Hadri, K., R. Larsson, and Y. Rao. 2012. Testing for stationarity with a break in panels where the time dimension is finite. *Bulletin of Economic Research* 64: 123–148.
- Harris, R.D.F., and E. Tzavalis. 1999. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91: 201–226.
- Hidalgo, J., and M. Schafgans. 2017. Inference and testing breaks in large dynamic panels with strong cross sectional dependence. *Journal of Econometrics* 196: 259–274.

- Karavias, Y., and E. Tzavalis. 2014. Testing for unit roots in short panels allowing for a structural break. *Computational Statistics & Data Analysis* 76: 391–407.
- . 2017. Local power of panel unit root tests allowing for structural breaks. *Econometric Reviews* 36: 1123–1156.
- . 2019. Generalized fixed-T panel unit root tests. *Scandinavian Journal of Statistics* 46: 1227–1251.
- Karavias, Y., P. Narayan, and J. Westerlund. 2021. *Structural breaks in interactive effects panels and the stock market reaction to COVID-19*. Mimeo.
- Kim, D. 2011. Estimating a common deterministic time trend break in large panels with cross sectional dependence. *Journal of Econometrics* 164: 310–330.
- . 2014. Common local breaks in time trends for large panels. *The Econometrics Journal* 17: 301–337.
- Kramer, W., W. Ploberger, and R. Alt. 1988. Testing for structural change in dynamic models. *Econometrica* 56: 1355–1369.
- Li, D., J. Qian, and L. Su. 2016. Panel data models with interactive fixed effects and multiple structural break. *Journal of the American Statistical Association* 111: 1804–1819.
- Liu, L., H.R. Moon, and F. Schorfheide. 2020. Forecasting with dynamic panel data models. *Econometrica* 88: 171–201.
- Makwasha, T., J. Wright, and P. Silvapulle. 2019. Panel data analysis of multi-factor capital asset pricing models. *Applied Economics* 51 (60): 6459–6475.
- Meltzer, A.H. 1960. Mercantile credit, monetary policy, and size of firms. *Review of Economics and Statistics* 42: 429–437.
- Nadiri, M.I. 1969. The determinants of trade credit in the US total manufacturing sector. *Econometrica, Journal of the Econometric Society* 37: 408–423.
- Okui, R., and W. Wang. 2020. Heterogeneous structural breaks in panel data models. *Journal of Econometrics*. Elsevier, 220 (2): 447–473.
- Pesaran, M.H. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967–1012.
- Pesaran, M.H., and A. Timmermann. 2004. How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting* 20 (3): 411–425.
- Petersen, M.A., and R.G. Rajan. 1997. Trade credit: Theories and evidence. *The Review of Financial Studies* 10: 661–691.
- Sarafidis, V., and T. Wansbeek. 2012. Cross-sectional dependence in panel data analysis. *Econometric Reviews* 31 (5): 483–531.
- Westerlund, J. 2006. Testing for panel cointegration with multiple structural breaks. *Oxford Bulletin of Economics and Statistics* 68: 101–132.
- Westerlund, J., H. Karabiyik, and P. Narayan. 2017. Testing for predictability in panels with general predictors. *Journal of Applied Econometrics* 32: 554–574.
- Wintoki, M.B., J.S. Linck, and J.M. Netter. 2012. Endogeneity and the dynamics of internal corporate governance. *Journal of Financial Economics* 105: 581–606.
- Zhu, H., V. Sarafidis, and M.J. Silvapulle. 2020. A new structural break test for panels with common factors. *The Econometrics Journal* 23 (1): 137–155.



More on Equilibrium Credit Rationing and Interest Rates: A Theory with New Evidence

94

Ying Wu

Abstract

In contrast to conventional understanding, an integrated theory of equilibrium credit rationing that allows restrictions on loan size as well as loan denials reveals a *positive* relationship between credit availability and interest rates when rationing by loan size (Gray, J. A. and Y. Wu, *Journal of Macroeconomics* 17: 405–420, 1995). This chapter uses the US bank lending practices from 1997 to 2010 as a quasi-natural experiment to test the interest rate hypothesis under loan-size rationing. Based on the vector error correction analysis of commercial and industrial loans made by the US domestic banks to large and middle-market firms, there is a cointegrating relation that sufficient tightness of credit market conditions (either an average loan size less than \$888,880 or a net percentage of banks tightening loan standards greater than -11%) rescues a positive relationship between loan sizes and loan rates. The net percentage of tightening banks plays a pivotal role in interacting with the size of loans and the demand for loans to maintain such a long-run equilibrium relationship.

Keywords

Equilibrium credit rationing · Interest rate · Loan size · Cointegrating relation

JEL Classification

E40 · E50

Y. Wu (✉)

Department of Economics and Finance, Franklin P. Perdue School of Business, Salisbury University, Salisbury, MD, USA
e-mail: yxwu@salisbury.edu

94.1 Introduction

Credit rationing, under which borrowers fail to get their desired funds even if they willingly pay higher interest rates, can exist as an equilibrium phenomenon in a completely free market in which there are no institutional and legal restrictions on interest rate adjustment. Asymmetric information is at the heart of equilibrium credit rationing. The well-known classical paper on equilibrium credit rationing by Stiglitz and Weiss (1981) posits information asymmetry in the form of *ex ante* private information about the riskiness of a firm's investment projects. Although the firm knows its own risk type, the intermediary that extends loans has no knowledge of the firm's risk type prior to making loans. Williamson (1987), on the other hand, analyzes the *ex post* private information about the firm's realized return on investment in a credit rationing model. The financial intermediary has to pay a monitoring cost to observe the outcome of projects that the firm can observe costlessly. In both of these studies, equilibrium credit rationing takes the form of restricting the number of loans whereby some borrowers get loans and others do not and thus rationing through restriction on loan size is not investigated at all. Consequently, changes in loan supply may be accompanied by no change at all in interest rates. While such a prediction of invariance of interest rates with respect to credit availability makes macroeconomists particularly intrigued, it is also interesting because it may suggest a new pattern of monetary policy transmission mechanism in terms of credit channel and interest rate behavior.

Besides the type of credit rationing by the number of loans studied by Stiglitz and Weiss, the credit-rationing literature has been exploring the other non-price types of credit rationing, which expands the scope of rationing and enriches the insights of the issue (see Kjenstad and Su 2015; Su and Zhang 2017). Nevertheless, while there are some studies that permit credit rationing by loan size, they either ignore the interest rate-credit relationship or take for granted that interest rates are negatively related to credit supply. Following the precursors to the equilibrium credit rationing literature,¹ Jaffee and Russell (1976) concentrates on credit rationing by size under *ex ante* asymmetric information but leaves the interest rate behavior under the rationing aside. Keeton (1979) studies credit rationing by both number and size of loans but it is not done in a consolidated setting: the possibility of rationing by loan number relies on either an interest rate ceiling or moral hazard in the lending process, neither of which is present in the model in which rationing by loan size occurs. Hence, it is impossible to evaluate the relative desirability of the two types of rationing. In Schreft and Villamil (1992), one of few efforts in exploring equilibrium credit rationing by loan size and interest rate behavior under the rationing, credit rationing occurs as a result of both the lender's market power (imperfect interest-rate discrimination) and imperfect information regarding the borrowers' characteristics, rather than depending on default risk and loan administration costs. Consequently,

¹Hodgman (1960), Freimer and Gordon (1962), Miller (1962), and Jaffee and Modigliani (1969), for examples.

instead of reflecting an endogenous effect on interest rates of changes in credit supply, the inverse relationship between loan sizes offered and interest rate charged plays a role only as a lender's strategy inducing borrowers to reveal their preferences for loan contract terms.

How do interest rates respond to tighter availability of loanable funds if credit rationing restricts the size rather than the number of loans? What new evidence concerning the relationship between credit supply and interest rates is available since the late 1990s after the two decades of intensive theoretical development of equilibrium credit rationing? For the first question, Gray and Wu (1995) provides a proof that tighter credit can cause interest rates to fall. In this chapter, I will first sketch an integrated theoretical framework to highlight the earlier findings for the first research question, and then I will turn to a vector-error correction model to address the second question above by analyzing the recent evidence of US banking practices and associated interest rate behavior with respect to credit supply.

94.2 A Theoretical Framework²

94.2.1 Model

Instead of appealing to any institutional or legal restrictions on interest rate adjustment or imperfect competition system, my study of loan-size rationing considers a competitive economy with a large number of identical households (ultimate savers/lenders), banks (intermediaries), and firms (ultimate investors/borrowers). In such a free-market economy flawed only by the *ex post* asymmetric information about borrowers' investment return, I focus on a representative member of each type of economic agents.³

Suppose that $\theta f(H)$ is the Cobb-Douglas production function with a uniformly distributed random term θ on the interval $[0, \bar{\theta}]$, H is the amount of the capital input employed by the firm and funded by bank loans. With ω denoting the capital rental rate and R the loan rate, the firm's profit can be expressed as $\theta f(H) - (1 + R)\omega H$. Although the debt contract specifies a fixed, non-contingent loan repayment at the end of the production period that the intermediary must collect from the firm, the firm may default if the value of production fails to cover the specified payment due to the adverse realization of random productivity parameter θ . The minimum value of the productivity factor θ required for the profit-maximizing firm's solvency, i.e., the threshold value of loan default θ^* , increases in both the loan rate and the amount of loans:

²The theoretical framework in this section is based on portions of Gray and Wu (1995) and Wu (2017).

³This setting naturally excludes the possibility for lenders to exercise price discrimination against borrowers, which helps, though, generate loan-size rationing in Schreft and Villamil (1992).

$$\theta^* = \frac{(1+R)\omega H}{f(H)}, \quad 0 < \theta^* < \bar{\theta} \quad (1)$$

Thus, the firm must choose H to maximize its expected profit

$$\int_{\theta^*}^{\bar{\theta}} \frac{[\theta f(H) - (1+R)\omega H]}{\bar{\theta}} d\theta \quad (2)$$

The firm's inverse input demand function for capital goods is derived from Eq. (2) as⁴

$$F(H) \equiv \frac{(\bar{\theta}/2)f'(H)}{[1 - (\varepsilon(H)/2)]} = 1 + R \quad (3)$$

where $0 < \varepsilon(H) < 1$ is the elasticity of output with respect to capital input. It can be shown that $F(H)$ is a decreasing function of loan size, H . If the intermediary honors the firm's inverse input demand so that the contract pair of loan size and loan rate exactly traces $F(H)$, then the $F(H)$ further generates the corresponding expected return to the intermediary, $\pi^F(H)$, which possesses a unique maximum.

On the other hand, the household endowed with a fixed amount of the capital good chooses to deposit more with the intermediary if the gross deposit rate, $1 + r$, gets higher⁵:

$$h = h(1 + r), \quad h' > 0 \quad (4)$$

In channeling purchasing power between households and firms, the intermediary chooses the loan rate to maximize its expected return on each "real" dollar loaned to the firm:

$$\max_{1+R} \left\{ \int_{\theta^*}^{\bar{\theta}} \frac{[(1+R)\omega H]}{\bar{\theta}} d\theta + \int_0^{\theta^*} \frac{[\theta f(H) - \gamma]}{\bar{\theta}} d\theta \right\} \frac{1}{\omega H} \quad (5)$$

where γ is the monitoring cost for the intermediary when the firm defaults. An increase in the loan rate affects the intermediary's expected return in two opposing ways. It raises the return by raising the loan payment when the firm is solvent but lowers the return by raising the probability of loan default. It is the second effect that produces the possibility of credit rationing.

⁴For simplicity, the rental price of capital, ω , is normalized at unity.

⁵Although the return the intermediary receives on a particular loan is random, the return on the intermediary's portfolio of loans to a large number of borrowers is not random. This assumption allows the intermediary to offer the households a nonrandom return on their deposits.

94.2.2 Rationing by the Number of Loans

The concavity of $\pi^F(H)$ with a turning point holds the key for credit rationing by the number of loans. In a competitive market, it is assumed that the intermediary will hold no excess reserves and loan out all its deposits; rationing loans by number means a provision of the loan size that maximizes the intermediary's return for those who receive loans while other rationed firms receive none. Whether or not the intermediary chooses to engage in number rationing depends on how plentiful the supply of credit is. If the supply of credit is sufficient to provide each loan applicant with a loan size at which the intermediary's return is maximized, the rationing by number will not occur because the intermediary would otherwise have to lower its return from each "real" dollar loaned by raising the loan size for some while rationing others. In this case, the relationship between interest rates and credit supply is negative, as traditionally expected. However, if the supply of credit is too tight to provide each firm a loan with the size that maximize the intermediary's return, the intermediary has incentive to engage in rationing by the number of loans. As a result, the level of interest rates is essentially anchored at the intermediary's optimal loan size; it is the number of loans rather than interest rates that comoves with the credit supply.

94.2.3 Rationing by the Size of Loans

Credit rationing by loan size differs from number rationing because the intermediary may no longer choose to deny loan to some firms but may stipulate, in addition to the loan rate, a ceiling of loan size to prevent firms from getting the desired amount of loans at the announced loan rate. The intermediary chooses the gross loan rate $(1 + R)$ for a loan of size H to maximize its expected return expressed in Eq. (5) and obtains its optimal loan schedule

$$I(H) \equiv \frac{\bar{\theta}f(H) - \gamma}{H} = 1 + R \quad (6)$$

The intermediary's optimal loan rate schedule further generates its schedule of expected return, $\pi^I(H)$, which can be shown to have a unique interior maximum. In addition, $\pi^I(H)$ is greater than $\pi^F(H)$ everywhere except at its tangency point with $\pi^F(H)$, H_C .⁶

As depicted in Fig. 1, when the loan size is as large as H_L , although the intermediary could raise its return by following its own loan schedule $I(H)$ and charging a loan rate of $1 + R_L^*$, the result is not attainable as the firm will not be forced to take the loan in excess of the amount that maximizes its profit. Therefore, for any loan sizes greater than H_C , it is $\pi^F(H)$ rather than $\pi^I(H)$ that determines the intermediary's expected return. When loan sizes are smaller than H_C , say H_T , the

⁶See Gray and Wu (1995) and Wu (2017) for proofs.

Becchetti et al. (2011) presents the robust evidence from Italy that the European Central Bank refinancing rate is significantly and positively related to credit rationing by loan size (partial credit rationing) but not by loan number (total credit rationing). As far as the type of credit rationing and its associated interest rate behavior are concerned, this finding is consistent with this present study has established in the previous section. However, the limitations in Becchetti et al. (2011) concern the type of interest rate, financial intermediaries, and borrowers in their dataset: it focuses exclusively on the central bank's policy rate rather than business loan rate, on the nonprofit-maximizing financial institutions rather than typical financial intermediaries as lenders, and on not-for-profit cooperatives rather than traditional for-profit industrial borrowers in the loanable funds market. In addition, Drakos and Giannakopoulos (2011) uses a micro dataset of approximately 9500 firms from 26 East European transition economies to classify credit-rationed firms by explicitly connecting between rationing and potential borrower profiles. While the survey results and the analysis from an estimated bivariate probit model identify the evidence of credit rationing by number, the model design leaves rationing by size uninvestigated. Using loan application and loan contract information from a large sample of Bulgaria, Kirschenmann (2016) measures credit rationing by the ratio of granted to requested loan amounts and provides direct evidence on the presence of loan size rationing as well as its extent and evolution over repeated interactions between small firms and their bank.

Based on an international dataset containing 18,000 business loans, Berger et al. (2018) provides their findings concerning credit rationing and interest rates in a period covering both before and during 2008–2009 financial crisis. They find that US banks appear to engage in more quantity rationing than banks from other nations and that the declines in loan quantities are larger for US banks than non-US banks and for US borrowers than non-US borrowers. Additionally, they find that as credit supply tightens, loan spreads tend to get larger for the public and relationship borrowers who experience size rationing while risky private and non-relationship borrowers could be denied from getting loans at all. While their results can still be consistent with the theoretical model developed in the Sect. 94.2 of this entry in terms of the relation between credit supply and interest rate behavior, the conceptual distinction of credit rationing by banks from credit-supply tightening in market appears ambiguous in Berger et al. (2018). More recently, employing microlevel data on small- and medium-sized enterprises collected by the European Central Bank over the period 2009–2019, Beyhaghi et al. (2020) further demonstrates not only that the equilibrium credit rationing can exist even in the presence of contract heterogeneity but also that rationing by the size rather than number of loans is most likely rationing regime in the post-financial crisis era.

94.4 Vector-Error Correction Model (1997–2010)

This section is intended to examine an extended sample period of US banking from 1997 to 2010 that covers two most recent economic recessions (the early 2000s recession from March 2001 to December 2001 and the Great Recession from

December 2007 to June 2009). With tight credit supply, the episodes of economic recession can be used as a quasi-natural experiment to test the theory of equilibrium credit rationing developed in this study, particularly the loan interest rate behavior and its relationship with loan size and credit-market conditions.

The quarterly dataset used in this study includes four major variables. Among them are the weighted-average effective loan rate (*LOANRATE*) and the average loan size (*LOANSIZE*) for all commercial and industrial loans made by all commercial banks, the net percentage of domestic banks tightening standards for commercial and industrial loans to large and middle-market firms (*TIGHTEN*), and the net percentage of domestic banks reporting stronger demand for commercial and industrial loans from large and middle-market firms (*LOAN_SD*). All these series are quarterly observations extracted from Federal Reserve Economic Database (FRED) and the Senior Loan Officer Opinion Survey on Bank Lending Practices.

In addition, an interaction term (*TIGHTEN_SIZE*) is added in order to investigate the composite impacts of loan size and tightened loan standard on the loan interest rate. Here, the effect of one variable (say, loan size) that forms the interaction depends on the level of the other variable (i.e., the net percentage of banks that tighten loans) in the interaction.

The methodology used to analyze the data is to estimate a vector-error correction model that helps understand the long-run equilibrium relationship (cointegration) among the business interest rate on loans, the size of loans, and credit market conditions. The precondition for the validity of cointegration estimation and vector-error correction analysis requires that the investigated series are nonstationary in their levels (I(1) process) but stationary in their first-order differences (I(0) process); if these conditions are met, there is possibly a linear combination of the series that is stationary, which is the so-called cointegration relation. Table 1 lists the results from three distinctive unit root tests of the five series for the cases of including intercept only and both intercept and trend, respectively. All the series satisfy the stationarity

Table 1 Unit-root test (1997Q2-2010Q4)

Info Rule	Test Type					
	Levin, Lin, and Chu		Im, Pesaran, and Shin		Fisher – ADF	
	Intercept only	Both intercept and trend	Intercept only	Both intercept and trend	Intercept only	Both intercept and trend
AIC	0.242	0.388	0.017	0.040	0.043	0.072
	0.000	0.000	0.000	0.000	0.000	0.000
SC	0.417	0.698	0.068	0.500	0.158	0.688
	0.000	0.000	0.000	0.000	0.000	0.000
HQ	0.242	0.477	0.017	0.236	0.043	0.399
	0.000	0.000	0.000	0.000	0.000	0.000

Notes

1. For information rules, AIC stands for Akaike Information Criterion, SC Schwarz Information Criterion, and HQ Hannan-Quinn Criterion
2. The first row for each and every information rule reports the *p*-value for the unit-root tests of the group variables at their levels, and the second row the *p*-value for the unit-root tests of the same group variables in their first-order differences

conditions for cointegration analysis. Next, the Johansen cointegration test is used to determine the number and properties of the cointegration relationship; the optimal lag length used in the test is one-quarter lag according to the Schwarz information criterion and Hannan-Quinn information criterion for the estimated vector autoregressions. As Table 2 shows, both the Johansen trace test and the maximum eigenvalue test jointly support that there is a cointegration relationship at 5% significance level conditional on the assumption that the level series and the cointegrating equations have linear trends.⁷

The vector-error correction model to be estimated can be presented as the following:

$$\begin{pmatrix} \Delta \text{LOANRATE}_t \\ \Delta \text{LOANSIZE}_t \\ \Delta \text{TIGHTEN}_t \\ \Delta \text{TIGHTEN_SIZE}_t \\ \Delta \text{LOAN_SD}_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{pmatrix} + (\beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5 \ \beta_6 \ \beta_7) \begin{pmatrix} \text{LOANRATE}_{t-1} \\ \text{LOANSIZE}_{t-1} \\ \text{TIGHTEN}_{t-1} \\ \text{TIGHTEN_SIZE}_{t-1} \\ \text{LOAN_SD}_{t-1} \\ \text{TREND} \\ 1 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{16} \\ \vdots & \ddots & \vdots \\ \gamma_{51} & \cdots & \gamma_{56} \end{pmatrix} \begin{pmatrix} \Delta \text{LOANRATE}_{t-1} \\ \Delta \text{LOANSIZE}_{t-1} \\ \Delta \text{TIGHTEN}_{t-1} \\ \Delta \text{TIGHTEN_SIZE}_{t-1} \\ \Delta \text{LOAN_SD}_{t-1} \\ 1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix} \quad (7)$$

All the variables in the five-equation system (7), including both the first-order difference terms and the cointegration equation, are stationary. The first part on the

⁷To get this result, the Pantula principle is used to test the *joint* hypothesis of both the cointegration rank order and the deterministic components in the multivariate vector error correction model. Specifically, the test is run continuously by moving through from the most restrictive alternative ($r = 0$ and the model in which the level data have no deterministic trends and the cointegration equations have intercepts) to the least restrictive alternative ($r = n - 1$ and the model in which the level data and the cointegration equations have linear trends). At each stage, the $\lambda_{\max}$ (or λ_{trace}) test statistic is compared to its critical value and the procedure continues until the null hypothesis is not rejected. For details, see Johansen (1992).

Table 2 Johansen cointegration test (1997Q2-2010Q4)

	Eigenvalue	Null hypothesis	Alter. hypothesis	Statistic	5% C.V.	p-Value
λ_{trace} test	0.629	$r = 0^*$	$r > 0$	128.324	88.803	0.000
	0.462	$r \leq 1^*$	$r > 1$	77.732	63.876	0.002
	0.339	$r \leq 2^*$	$r > 2$	46.058	42.915	0.023
	0.262	$r \leq 3$	$r > 3$	24.901	25.872	0.065
λ_{max} test	0.629	$r = 0^*$	$r = 1$	50.592	38.331	0.001
	0.462	$r = 1$	$r = 2$	31.673	32.118	0.056
	0.339	$r = 2$	$r = 3$	21.156	25.823	0.183
	0.262	$r = 3$	$r = 4$	15.545	19.387	0.165

Notes

1. r in the table represents the number of cointegration equations. Both trace test and Max-eigenvalue test indicate two cointegrating equations at the 0.05 level
2. * denotes rejection of the hypothesis at the 5% significance level

right-hand side of Eq. (7) gives the response of the dynamic system to disturbances that let the system deviate from its long-run equilibrium. Let the vector β be a 1×7 row cointegrating vector, $y_t - 1$ the 7×1 column vector of all the variables in the cointegration equation, and α a 5×1 column vector. If $\beta y_t - 1 = 0$, the system is in its long-run equilibrium; if $\beta y_t - 1 \neq 0$, that is, the system exhibits a deviation from the equilibrium, then $\alpha \beta y_t - 1$ represents the error correction that helps move the system toward its equilibrium. The second part on the right-hand side of Eq. (7) is the traditional vector autoregression system in first differences.

The estimated cointegration equation reveals a robust long-run equilibrium relationship that provides fresh evidence for the theory of credit rationing by loan size developed in the Sect. 94.2 of this chapter. Specifically, as reported in Table 3, a decrease of an average thousand dollar in commercial and industrial loans *lowers* rather than raises the weighted average effective loan rate by 1.4 basis point.⁸ In addition, each net percentage point increase in the domestic banks tightening standards for commercial and industrial loans to large and middle-market firms *lowers* rather than raises the weighted average effective loan rate by 80 basis points. According to these two estimates, tightened loan standard by commercial banks and reduced average loan size, instead of elevating interest rates on loans (as traditionally construed) or leaving them invariant (as predicted by many equilibrium credit rationing models that focus only on rationing by loan numbers), are respectively associated with *lower* loan interest rates. In addition, each net percentage point increase in the banks reporting stronger demand for loans from firms is associated with a 25 basis-point decrease in the loan rate, as presented in the estimated cointegration equation, which further supports the theory of equilibrium credit rationing by loan size, because tight credit market conditions represent large *net* demand for credit (i.e., either strong demand for or weak supply of loans or both).

⁸One basis point is 1/100 percentage point.

Table 3 Vector error correction estimates

Cointegration equation:

$$\begin{aligned} \text{LOANRATE} = & -0.601 + 0.014\text{LOANSIZE} - 0.808\text{TIGHTEN} + 0.001\text{TIGHTEN_SIZE} \\ & (3.185) \quad (-5.870) \quad (4.878) \\ & - 0.250\text{LOAN_SD} - 0.092\text{TREND}(97\text{Q}2) \\ & (-7.599) \quad (-2.469) \end{aligned}$$

Cointegration equation (CointEq1)					
LOANRATE(-1)	1.000				
LOANSIZE(-1)	-0.014***				
TIGHTEN(-1)	0.808***				
TIGHTEN_SIZE(-1)	-0.001***				
LOAN_SD(-1)	0.250***				
TREND(97Q2)	0.092***				
Constant	0.601				
Dependent variable	D(LOANRATE)	D(LOANSIZE)	D(TIGHTEN)	D(TIGHTEN_SIZE)	D(LOAN_SD)
Error correction and VAR terms					
CointEq1	-0.018 (-1.001)	12.046*** (4.032)	1.197*** (2.415)	1238.057*** (3.289)	-1.880*** (-2.990)
D(LOANRATE(-1))	0.565*** (4.349)	-7.805 (-0.378)	-3.840 (-1.121)	-2853.862 (-1.097)	5.345 (1.230)
D(LOANSIZE(-1))	-0.000 (-0.001)	-0.248** (-1.919)	0.021 (0.985)	17.403 (1.067)	-0.027 (-0.994)
D(TIGHTEN(-1))	0.026 (1.505)	-5.583** (-2.005)	-0.828** (-1.793)	-612.372** (-1.746)	1.108** (1.891)
D(TIGHTEN_SIZE(-1))	-0.000 (-1.703)	0.005* (1.644)	0.001* (1.656)	0.556 (1.279)	-0.001** (-1.665)
D(LOAN_SD(-1))	0.005 (1.145)	-1.376** (-1.831)	-0.405*** (-3.250)	-270.064*** (-2.850)	0.176 (1.111)

Notes: The estimates with *** are significant at 1% level,
The estimates with ** are significant at 5% level,
The estimates with * are significant at 10% level

Given the significance of *LOANSIZE* and *TIGHTEN* in the cointegration equation, their interaction term gives additional insight of how each of them affects *LOANRATE*. On one hand, the effect of tightened loan standard on the loan rate depends on the magnitude of loan size as $\frac{\partial(\text{LOANRATE})}{\partial(\text{TIGHTEN})} = -0.8 + 0.001(\text{LOANSIZE})$, implying that banks engaged in tightening loans will reduce the loan rate to lower credit risk if the average loan size is less than \$888,880 (see Panel a of Fig. 2). On the other hand, the effect of loan size on the loan rate depends on how large the net percentage of the domestic banks tightening standards for commercial and industrial loans is. Given $\frac{\partial(\text{LOANRATE})}{\partial(\text{LOANSIZE})} = 0.01 + 0.001(\text{TIGHTEN})$, it follows that smaller loan size is associated with lower loan rates if and only if net percentage of the domestic banks loosening standards is less than 11% (i.e., those tightening standards greater than -11%, as shown in Panel b of Fig. 2). That is, it is the sufficient tightness of credit market conditions, even though the percentage of banks tightening loans falls short of the percentage of banks loosening loans when the net percentage of banks tightening is negative, that rescues a positive relationship between loan size and loan rate. These results from the interactive term analysis are well in line with not only the theory of equilibrium credit

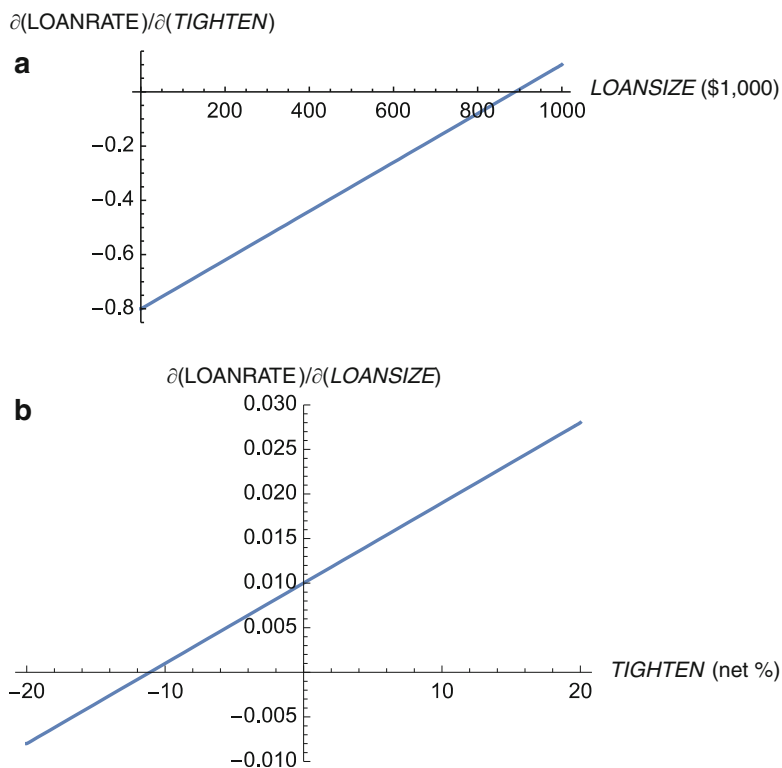


Fig. 2 The interacting impact of loan size and tightening banks. (Panel **a**) The loan-size impact on $\frac{\partial(\text{LOANRATE})}{\partial(\text{TIGHTEN})}$. (Panel **b**) The Interacting Impact of Loan Size and Tightened Loan Standard on $\frac{\partial(\text{LOANRATE})}{\partial(\text{LOANSIZE})}$

rationing by restricting loan size in Sect. 94.2 but also with the data for the sample period. Figure 3 presents the time plots of the US average loan sizes for commercial and industrial loans and the net percentage of the domestic commercial banks tightening standard of loans from 1997 to 2010. In the sample period, the loan size is uniformly below the threshold in favor of a negative relationship between credit tightening and loan rates (panel a of Fig. 3), and the most of the observed net percentage of tightening banks is well above the threshold in support of a positive relationship between loan sizes and loan rates (panel b of Fig. 3).

While Table 3 gives all the estimates of both the error-correction part and autoregression part of the entire vector error correction model, Tables 4 and 5 report the results of likelihood-ratio test to verify the significance of the estimated cointegrating vector β and the estimated speed-of-adjustment coefficient vector α . All the endogenous variables share the common cointegration trend and are in the cointegration space. In addition, in the cointegrated system, LOANRATE is not Granger caused by other variables at the 5% significance level (the first row of Table 6). Except that LOANRATE is weakly exogenous with respect to the cointegrating vector, all the other four endogenous variables are significantly

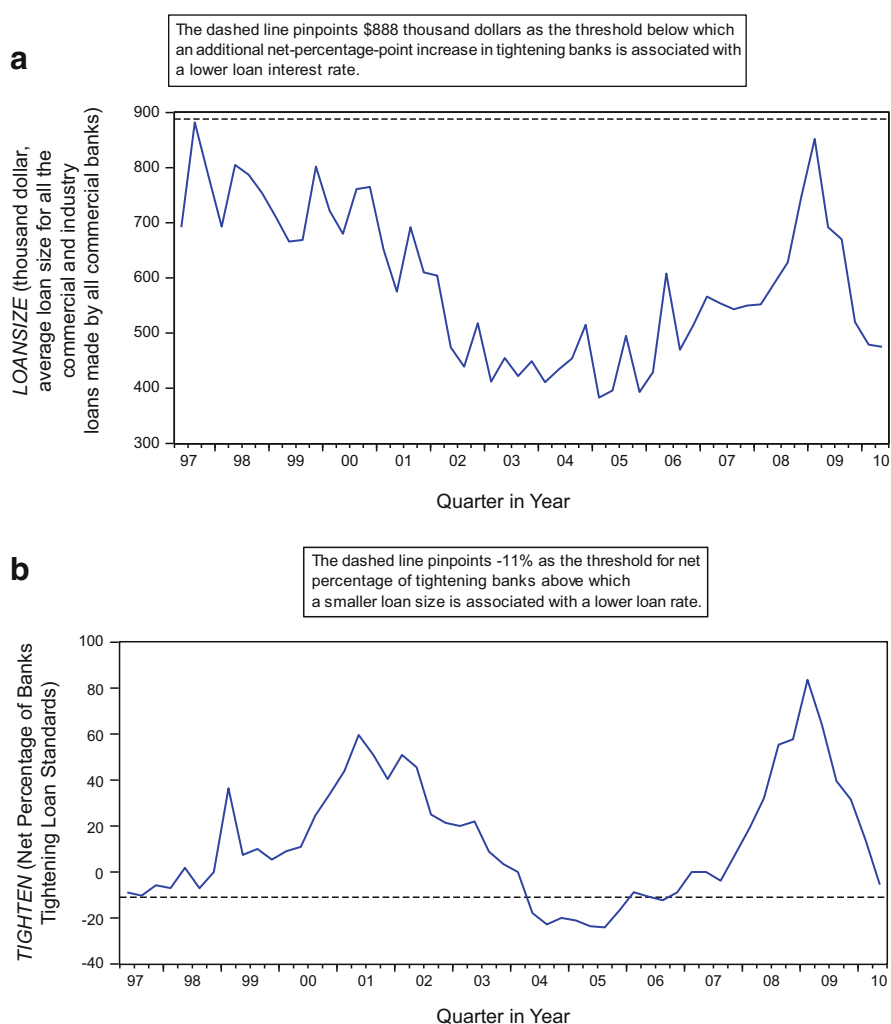


Fig. 3 Loan size and tightening banks in the USA (1997–2010). (Panel **a**) Average loan size for commercial and industry loans. (Panel **b**) Net percentage of banks tightening loan standards

endogenous in terms of their responsive adjustment to the cointegrating vector over time (long-run Granger causality).⁹ Indeed, the exogeneity result for the loan rate is consistent with the theoretical model of loan-size rationing. Once a bank finds it

⁹The evidence from a set of diagnostic tests supports the validity of the estimated vector error-correction model. For examples, the estimated dynamic system is stable, as all the inverse roots of the characteristic autoregression polynomial lie inside the unit circle and VEC specification imposes four roots on the unit circle; the Jarque-Bera residual normality test indicates that the estimated residuals are multivariate normal; and there is no serial correlation of residuals by the VEC residual serial correlation Lagrange multiplier test and portmanteau autocorrelation test.

Table 4 Significance test of variables in the cointegrating vector

Restrictions	$\beta_{\text{LOANRATE}} = 0$	$\beta_{\text{LOANSIZE}} = 0$	$\beta_{\text{TIGHTEN}} = 0$	$\beta_{\text{TIGHTEN SIZE}} = 0$	$\beta_{\text{LOAN SD}} = 0$
$\chi^2(1)$	3.670*	5.877**	10.317***	6.916***	17.729***
<i>p</i> -value	0.055	0.015	0.001	0.008	0.000

Notes: The estimates with *** are significant at 1% level, those with ** are significant at 5% level, and those with * are significant at 10% level

Table 5 Significance test of speed-of-adjustment coefficient to the cointegrating vector

Restrictions	$\alpha_{\text{LOANRATE}} = 0$	$\alpha_{\text{LOANSIZE}} = 0$	$\alpha_{\text{TIGHTEN}} = 0$	$\alpha_{\text{TIGHTEN SIZE}} = 0$	$\alpha_{\text{LOAN SD}} = 0$
$\chi^2(1)$	0.863	11.478***	4.344**	6.494***	4.668**
<i>p</i> -value	0.352	0.000	0.037	0.010	0.030

Notes: The estimates with *** are significant at 1% level, those with ** are significant at 5% level, and those with * are significant at 10% level

optimal to lower the loan rate, credit rationing on the loan size will take place and the loan standards will be tightened whereas the quantity of loans demanded is increasing at the lower loan rate.

Table 6 helps investigate the degree of endogeneity of variables and possible causality relations among them. The pairwise Granger causality test identifies the short-run causality among the variables in first differences by the significance of the pertinent coefficients (γ s), and the t-statistic for the lagged error-correction coefficient (α s) concerns the long-run causality in terms of how responsive a variable is to disequilibrium errors. In the short run, while a change in the loan size per se does not seem to cause others to respond significantly, its interaction term with the net percentage of banks tightening loan standard exhibits Granger-causality relations at the 10% level with all the other variables. Furthermore, as shown in the column entitled “Block Exogeneity,” the joint significance of all the included variables for short-run Granger causality holds only for the two dependent variables, i.e., net percentage of banks tightening loan standard and its interaction term with loan size, suggesting their strong endogeneity. Additionally, both the net percentage of banks tightening loan standard and the net percentage of banks reporting stronger loan demand Granger-cause all other variables except the loan rate. Indeed, as predicted by the theoretical model in Sect. 94.2, the loan rate either declines only with tightened credit rationing on loan size or gets anchored at its optimal level when credit rationing occurs by restricting the number of loans. In the long run, except for the loan rate, all the other variables play significant role (at the 1% level) in responding and adjusting to disequilibrium error and therefore moving to restore the long-run equilibrium relationship.

The findings from innovation accounting (impulse responses and variance decomposition) supports the results obtained from the cointegration relationship and short-run dynamic adjustment to disequilibrium errors. Figure 4 plots impulse responses based on the vector moving average representation of the expanded vector autoregression in standard form including the error-correction process, which exhibits the entire time paths of the five endogenous variables in response to the

Table 6 Short-run and long-run granger causality

	Excluded variables						t-Statistic for error correction term (speed-of-adjustment coefficient)
	χ^2 Statistics (for excluded variables)						
Dependent variables	D(LOANRATE)	D(LOANSIZE)	D(TIGHTEN)	D(TIGHTEN_SIZE)	D(LOAN_SD)	Block Exogeneity	
D(LOANRATE)	—	0.000	2.265	2.902*	1.312	3.714	−1.001
D(LOANSIZE)	0.142	—	4.023**	2.705*	3.352*	5.582	4.032***
D(TIGHTEN)	1.256	0.970	—	2.744*	10.567***	13.395***	2.415***
D (TIGHTEN_SIZE)	1.204	1.138	3.048*	—	8.127***	10.937**	3.289***
D(LOAN_SD)	1.512	0.989	3.576*	2.775*	—	6.574	−2.990***

Notes: The estimates with *** are significant at 1% level, those with ** are significant at 5% level, and those with * are significant at 10% level

white-noise disturbances in them respectively.¹⁰ In Fig. 4, the impulse response of LOANRATE over a period of 20 quarters to shocks on all the other four variables are consistent with what the cointegration equation tells about the long-run behavior of banks (the first row). As shown in the third row, the net percentage of banks tightening loans increases as both loan rate and loan size decrease while increasing with loan demand (the third row). In the fourth row, with *TIGHTEN_SIZE* interpreted as the average size of loans made by the net percentage of tightening banks, although its response to loan demand is positive, its impact on the loan rate is uniformly negative. These findings well echo the theoretical results of equilibrium credit rationing by loan size.

While impulse response presents the response of a variable to an impulse in another variable, the chart of variance decomposition reported in Table 7 shows how much a shock to one variable impacts the (variance of the) forecast error of a different variable. Over the entire period of as long as 20 quarters, the net percentage of tightening banks explains more than 90% of its own forecast error variance and the loan rate explains about 84% of its own forecast error variance, suggesting the strong exogeneity of these two series in the dynamic adjustment process. This can be explained by the loan-size rationing model in which loan-rate reduction and credit rationing are indeed the two sides of the same coin. In contrast, the interaction term (*TIGHTEN_SIZE*) is strongly endogenous; about 92% of its forecast error variance is explained by the net percentage of tightening banks. Meanwhile, the loan size also exhibits relatively strong endogeneity because both the net percentage of tightening banks and the demand for loans influence the forecast error variance of loan size by 35% and 15%, respectively. In addition, by the end of 20 quarters after the shocks, the shocks to the net percentage of tightening banks and loan size respectively account for 31% and 16% of the forecast error variance of loan demand, and their interaction term accounts for 42% of the forecast error variance of loan demand.

94.5 Conclusion

This study is a follow-up work to Gray and Wu (1995) with the fresh empirical evidence on the US market of commercial and industrial loans from 1997 to 2010. In the earlier theoretical framework, it was shown that the nature of the relationship between credit supply and interest rates is sensitive to the form that credit rationing takes. When credit rationing takes the form of restricting the number of loans, interest rates are insensitive to fluctuations in the credit supply. Nevertheless, the result no longer holds when it is the size rather than number of loans that is restricted to borrowers in the market for loanable funds. If rationing is allowed to take the form

¹⁰The Choleski decomposition with degree of adjustment is used in innovation accounting in order to identify the underlined structural model. Based on the matrix of pairwise cross-correlograms, the cross-correlation coefficients for the estimated residuals in the model are essentially within plus or minus two times the asymptotic standard errors of the lagged correlations. Therefore, the ordering of the variables in innovation accounting is unimportant.

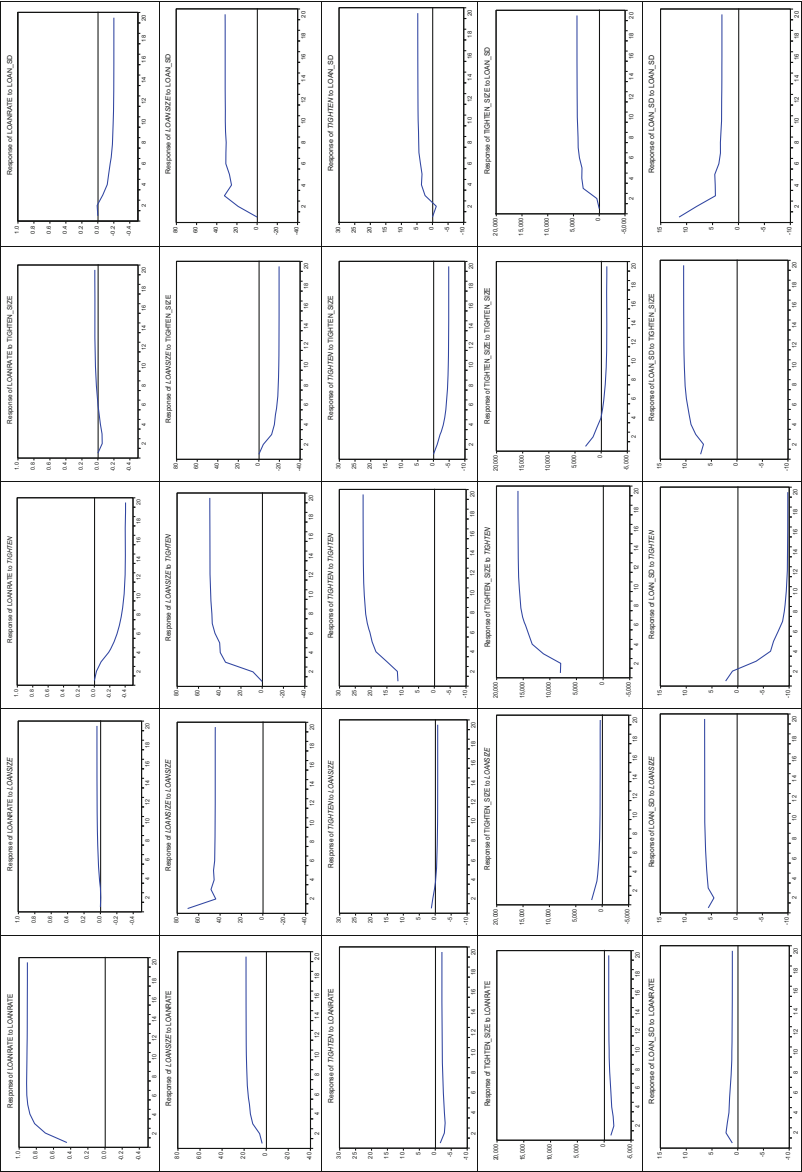


Fig. 4 Impulse responses

Table 7 Variance decomposition

Variance decomposition of LOANRATE							
Period	S.E.	LOANRATE	LOANSIZE	TIGHTEN	TIGHTEN_SIZE	LOANSZ	LOANSZ
1	0.443678	100.0000	0.000000	0.000000	0.000000		0.000000
5	1.754353	95.13553	0.020226	3.438453	0.226641		1.179154
10	2.829042	88.46217	0.080357	8.890128	0.113244		2.454101
15	3.621422	85.29830	0.113221	11.46526	0.118923		3.004294
20	4.270967	83.75240	0.129626	12.72328	0.126541		3.268155
Variance decomposition of LOANSIZE							
Period	S.E.	LOANRATE	LOANSIZE	TIGHTEN	TIGHTEN_SIZE	LOANSZ	LOANSZ
1	70.46057	0.272408	99.72759	0.000000	0.000000		0.000000
5	148.2384	2.882107	61.18728	20.16797	3.003391		12.75926
10	226.3612	4.186494	45.84151	30.35221	4.708545		14.91123
15	285.9942	4.644063	40.84734	33.76423	5.297376		15.44699
20	335.3749	4.860087	38.51360	35.35896	5.573553		15.69379
Variance decomposition of TIGHTEN							
Period	S.E.	LOANRATE	LOANSIZE	TIGHTEN	TIGHTEN_SIZE	LOANSZ	LOANSZ
1	11.69338	1.759624	1.071553	97.16882	0.000000		0.000000
5	36.36517	2.646811	0.169968	92.55833	2.263096		2.361797
10	62.19743	1.615504	0.112935	91.63325	3.359480		3.278834
15	81.47399	1.276901	0.111092	91.32703	3.713750		3.571231
20	97.11161	1.127698	0.111180	91.19284	3.869966		3.698315

of restrictions on loan size, the tight credit supply can cause banks to lower rather than raise interest rates (both the loan rate and deposit rate) when size rationing occurs in order to reduce default risk. Furthermore, even when credit is not tight enough to cause size rationing to occur, reduced supply of loans may still be associated with decreases in deposit rates. Only when credit availability is loose can a traditional negative relationship between loan supply and interest rates be honored.

The contribution of the present entry is to use a data sample of 1997–2010 US bank lending practices as a quasi-natural experiment to test the theory of equilibrium credit rationing, focusing on the loan interest rate behavior and its relationship with loan size and credit-market conditions. The vector error correction analysis finds a robust positive relationship between loan-size rationing and loan interest rates for a long-run equilibrium in the market for commercial and industrial loans. Examining the composite impacts of loan size and tightened loan standard on the loan interest rate for the sample period further suggests that it is sufficient tightness of credit market conditions that rescues a positive relationship between loan size and loan rate, especially when the economy experiences recessions. Specifically, on one hand, credit tightening lowers the loan interest rate if the loan size is less than \$888,880, and on the other hand, a smaller loan size is associated with lower loan interest rates if the net percentage of banks tightening loan standard is greater than -11% . According to Granger causality and innovation accounting analyses, the net percentage of tightening banks plays a pivotal role in interacting with loan size and loan demand to maintain the long-run equilibrium positive relationship between credit supply and interest rates.

94.6 Cross-References

- [Equilibrium Credit Rationing and Monetary Non-neutrality in a Small Open Economy](#)

References

- Becchetti, L., M.M. Garcia, and G. Trovato. 2011. Credit rationing and credit view: Empirical evidence from an ethical bank in Italy. *Journal of Money, Credit and Banking* 43 (6): 1217–1245.
- Berger, A.N., and G.F. Udell. 1992. Some evidence on the empirical significance of credit rationing. *Journal of Political Economy* 100: 1047–1077.
- Berger, A. N., T. Makaew, and R. T. Ariss. 2018. Who pays for financial crises? price and quantity rationing of different borrowers by domestic and foreign banks. *IMF Working Paper* No. 18/158.
- Beyhaghi, M., F. Firoozi, A. Jalilvand, and L. Samarbakhsh. 2020. Components of credit rationing. *Journal of Financial Stability* 50: 1–14.
- Drakos, K., and N. Giannakopoulos. 2011. On the determinants of credit rationing: Firm-level evidence from transition countries. *Journal of International Money and Finance* 30 (8): 1773–1790.

- Evans, D., and B. Jovanovic. 1989. An estimated model of entrepreneurial choice and liquidity constraints. *Journal of Political Economy* 97 (4): 808–827.
- Freimer, M., and M. Gordon. 1962. Why banks ration credit. *Quarterly Journal of Economics* 79: 397–410.
- Gray, J.A., and Y. Wu. 1995. On equilibrium credit rationing and interest rates. *Journal of Macroeconomics* 17 (3): 405–420.
- Hodgman, D. R. 1960. Credit risk and credit rationing. *Quarterly Journal of Economics* 74: 258–278.
- Jaffee, D. M., and F. Modigliani. 1969. A theory and test of credit rationing. *American Economic Review* 59: 850–872.
- Jaffee, D.M., and T. Russell. 1976. Imperfect information, uncertainty, and credit rationing. *Quarterly Journal of Economics* 90 (4): 651–666.
- Johansen, S. 1992. Determination of cointegration rank in the presence of a linear trend. *Oxford Bulletin of Economics and Statistics* 54: 383–397.
- Keeton, W.R. 1979. *Equilibrium credit rationing*. New York/London: Garland Publishing.
- Kirschenmann, K. 2016. Credit rationing in small firm-bank relationships. *Journal of Financial Intermediation* 26: 68–99.
- Kjenstad, E., and X. Su. 2015. Credit rationing by loan size: A synthesized model. *The Quarterly Review of Economics and Finance* 55: 20–27.
- Miller, M. 1962. Credit risk and credit rationing: further comment. *Quarterly Journal of Economics* 76: 480–488.
- Schreft, S. L., and A. P. Villamil. 1992. Credit rationing by loan size in commercial loan markets. *Economic Review* 78, Federal Reserve Bank of Richmond: 3–7.
- Stiglitz, J.E., and A. Weiss. 1981. Credit rationing in markets with imperfect information. *American Economic Review* 71: 393–410.
- Su, X., and L. Zhang. 2017. A reexamination of credit rationing in the Stiglitz and Weiss model. *Journal of Money, Credit and Banking* 49 (5): 1059–1072.
- Williamson, S.D. 1987. Costly monitoring, loan contracts, and equilibrium credit rationing. *Quarterly Journal of Economics* 102: 135–145.
- Wu, Y. 2017. *An analysis of credit and equilibrium credit rationing*, Vol. 10, Routledge Library Editions: Monetary Economics. London/New York: Routledge (an earlier edition was published in 1994 by New York and London: Garland Publishing, Inc., which was listed in “Annotated Listing of New Books,” *Journal of Economic Literature* 32(4): 1966).



The Effect of Basel III on Banks' Lending

95

Hao Chang and Yan Xue

Abstract

Basel III is considered as the most comprehensive set of regulations for mitigating banks' credit risk and market risk. Its major function is to encourage banks to raise more capital and carry sufficient liquid assets to avoid bank runs and defend financial crisis, thus promoting the stability of the financial system and maximizing the social welfare. Although stricter capital and liquidity requirements reduce the risk of bank runs, a voice of opposition about the negative impact of the regulations on banks' lending and profitability exists all the time. Finance researchers have conducted a number of theoretical and empirical studies to investigate how the rigorous regulations affect banks' risk and performance. This chapter summarizes this growing literature and sheds light on the future research areas on how capital and liquidity requirements in Basel III effect on Banks' lending and performance.

Keywords

Basel III · Capital regulations · Liquidity regulations · Banks' lending · Risk management

Hao Chang is a Quantitative Research Analyst at Jacobs Levy Equity Management. The views expressed in this article are solely those of the authors, and do not necessarily reflect the views of Jacobs Levy Equity Management.

H. Chang
Jacobs Levy Equity Management, Florham Park, NJ, USA

Y. Xue (✉)
Rutgers University, Piscataway, NJ, USA
e-mail: x7y599chh@outlook.com

95.1 Introduction

The Basel Accords – a set of banking supervision regulations – were set by the Basel Committee on Banking Supervision (BCBS). The Basel Accords were developed to be the most comprehensive set of regulations for managing banks' credit risk and market risk. The key function of Basel is to encourage banks to absorb more capital and hold enough liquid assets to avoid bank runs and survive in a financial crisis. The regulations are designed to maintain the stability of the financial system by setting minimum capital and liquidity regulations on banks.

Although more stringent capital and liquidity regulations can reduce the risk of bank runs, there has been a vigorous debate about the potential negative impact of the regulations on bank lending and bank profitability. Many economists develop theoretical models or design empirical studies to explore how the tough regulations affect the risk the banks face and the banks' performance in the financial markets. A growing strand of literature discusses the impacts of capital requirements, liquidity requirements, and the joint effect of these two kinds of regulations. This chapter introduces corresponding studies and summarizes the effect of Basel III on Banks' lending.

The remainder of the chapter is organized as follows, Sect. 95.2 describes the historical evolution of Basel Accords. Section 95.3 investigates the impact of capital requirements on banks' lending. Section 95.4 discusses how the liquidity regulations affect banks' lending. Section 95.5 presents the interaction between the capital and liquidity regulations. Section 95.6 concludes.

95.2 From Basel I to Basel II

The Basel Accords are a series of banking regulation agreements. They were set by the Basel Committee on Bank Supervision (BCBS). The committee provides recommendations on banking and financial regulations about credit risk, liquidity risk, and operational risk. The Basel Accords were modified several times between 1980 and 2011 and can be broken down into Basel I, Basel II, and Basel III.

95.2.1 Basel I

Basel I, or the Basel Capital Accord, was formed in 1988. In 1980s, the number of international banks and large-size banks was growing faster and faster. The regulators realized that the failure of these banks would cause severe consequences in the economy since the cash reserves held by the banks were not sufficient. Besides that, the integration between banks increased rapidly. The deep integration might bring a crisis in multiple countries when one large bank failed in financial distress. Based on that fact, G10 countries enforced Basel I in 1992 and more than 100 countries also implemented these regulations.

Basel I is the first framework to help banks manage credit risk by governing the banks' asset allocations. The international banks are guided to allocate capital to lower-risk investments. Different kinds of assets are given different risk weights. Basel I classifies the assets into four categories based on their risk levels:

- Risk-free assets, like cash and treasury bonds (0% risk weight)
- Loans to other banks or securities (20% risk weight)
- Residential mortgages (50% risk weight)
- Corporate debts (100% risk weight)

Basel I requires a significant international bank to hold 8% of their risk-weighted assets as cash reserves and provides banks incentives to invest more in liquid securities and loans instead of risky mortgage and corporate bonds.

95.2.2 Basel II

Basel II was implemented in 2004 to extend Basel I. Regulators realized that the capital requirements in Basel I only considered credit risk while operational and market risks were ignored. As a result, Basel II strengthened the minimum capital requirements, supervisory mechanisms, and market discipline in Basel I. Banks are guided to determine their minimum capital requirements with the standardized measures of credit, operational, and market risks. Basel II attempts to create a more comprehensive framework to lower the risk of banking system.

A simple capital to asset ratio in Basel I might lead to more risky asset portfolios held by banks (Koehn and Santomero 1980; Kim and Santomero 1988). The Basel II framework makes more efforts to measure the risks of assets and introduces internal rating models to determine capital charges for individual credit risks.

In addition, there are also some key modifications in Basel II. First, it creates standardized measures for evaluating operational risk. Second, it uses market value instead of book value to calculate credit exposure. Finally, it considers the information asymmetry problem of banks. The disclosure requirements were developed to improve the supervisory mechanisms and market transparency.

95.2.3 Basel III

The Global Financial Crisis in 2008 exposed the weakness of the banking system. Therefore, Basel III was developed in November 2010. However, the tough requirements of Basel III led to a huge argument and they were not be fully implemented in the global banking systems.

The key reasons of the 2008 financial crisis include poor corporate governance, unsuccessful liquidity management, and over-leveraged capital structure.

Correspondingly, in Basel III, the minimum capital requirements were strengthened and the liquidity rules were designed to mitigate the short-term credit stress in a crisis. The liquidity rules include two key standards: the liquidity coverage ratio (LCR) and the net stable funding ratio (NSFR). Basel III requires banks to maintain the following financial ratios:

- Tier 1 capital ratio (Equity capital/Risk-weighted assets) $\geq 4.5\%$
- Leverage ratio (Tier 1 capital/Average total assets) $\geq 3\%$
- Liquidity coverage ratio (Liquid assets/Total outflows over next 30 days) $\geq 100\%$
- Net stable funding ratio (Available stable funding/Required stable funding) $\geq 100\%$

Most banks in important countries were required to comply with the new regulations between 2015 and 2019. Banks began to report the LCR in January 2015 and were fully binding in January 2019. The NSFR was fully implemented in January 2018. Since the LCR was introduced first, academia focuses more on the performance of LCR. Note that the LCR asks banks to maintain a buffer of high-quality liquid assets (HQLA) equal to or above 100% of total net cash outflows over the next 30-day calendar period. Banks report the LCR at least once a month. But the reporting frequency may be increased during a crisis. The HOLA is composed of the level 1 and level 2 assets. Level 1 assets comprise of cash, bank notes, highly liquid securities, and statutory reserves. Level 2 assets include high-grade corporate bonds, covered bonds and claims on quasi-government entities such as municipal bonds. The total net cash outflows are defined as either the bank's expected 30-day outflows less the greater of its expected 20-day inflows or 25% of its anticipated 30-day outflows.

The NSFR aims to ensure that banks maintain a stable funding structure. It is expressed as a ratio that must equal to or exceed 100%. The ratio relates the bank's available stable funding (ASF) to its required stable funding (RSF). A bank's total ASF is the portion of its capital and liabilities that will remain with the institution for more than 1 year. A bank's total RSF is the amount of stable funding that it is required to hold given the liquidity characteristics and residual maturities of its assets and the contingent liquidity risk arising from its off-balance sheet exposures. The NSFR became a minimum standard applicable to all internationally active banks on a consolidated basis on 1 January 2018, although national supervisors may also apply it to any subset of entities of large internationally active banks or to all other banks. Banks must meet the NSFR requirement on an ongoing basis and report on a quarterly basis.

Note that the US Federal Reserve exempts small banks from the Basel III liquidity requirements. The reason is that the burdens on small banks which come from the liquidity regulations outweigh the liquidity risk the small banks facing in a crisis. Besides that, the FDIC helps to resolve the distress small banks face. In contrast, the European Banking Authority requires banks of all sizes to comply with the required liquidity ratios. The European Banking Authority thinks that the credit supply of small banks is very important for the society.

In addition, Basel III includes a more complex and stringent reserve requirement and a countercyclical capital buffer to relax regulations during the financial distress.

95.3 Capital Regulation and Banks' Lending

Capital requirements are demonstrated to be effective tools to ensure the safety and soundness of banks. About 20 well-known professors once supported implementing the bank capital requirements and highlighted the important roles of capital requirements to avoid financial crises in an open letter to the Financial Times.

Although most of the economists agree with the important role of capital requirements, there is always a debate on the optimal capital requirements for banks. Some scholars prefer to loose capital requirements. They argue that a high required capital ratio would restrict banks' ability to provide credit to the economy, which reduces the consumption and economic growth. The cost of capital requirements can be reflected in the banking system: equity requirements would restrict lending and impede growth. A strand of the theoretical banking literature has argued that the tough requirements would reduce bank credit.

The capital requirements imply more equity with greater funding costs. The costliness of equity capitalization cannot be ignored. The changes in funding costs can be explained by Modigliani and Miller (1963). The banks have technological advantage in funding with deposits (Diamond and Rajan 2001). Comparing with the deposits, the equity is more expensive to adsorb. Besides the direct effect, the capital requirements raise costs by the tax subsidy of debt (Kashyap et al. 2010). In addition, the costs of equity capital could deviate from the government guarantee of bank debt (Kelly et al. 2016). Because of these types of additional costs, higher capital requirements lead banks to curtail their risk-weighted assets or lending to meet the standards.

However, Deli and Hasan (2017) argue the negative effect of capital requirements on banks' lending growth is weak. They link the effect of capital requirements with the banks' capital structure. The negative impact of capital requirements becomes low if banks are well-capitalized and disappears for banks with high capital ratios. According to their results, the cost of the capital requirements on bank's lending is not as high as expected.

Besides that, the impact of capital requirements may be tiny according to the Modigliani-Miller model. In response to higher capital requirement, the banks have to issue more equity to maintain the lending. But the deleveraging has a marginal impact on the cost of funding for banks. When the leverage become lower, the cost of equity decreases (Hanson et al. 2011).

Furthermore, some economists provide a counterintuitive suggestion that the tighter capital requirements help to increase lending. The benefits stem from two aspects. First, the capital requirements lower the probability of bank defaults in a crisis. Through increasing capital stringency, banks adjust their asset portfolio and decrease their risky investments. As a result, banks focus more on higher-quality lending and contribute to a healthy economic growth in a long run (Deli and Hasan

2017). Second, Begenau (2020) shows that more rigorous capital requirements can increase the banks' lending by lowering the deposit rate in a general equilibrium. The tough capital requirements thus help to reduce banks' funding costs and increase lending. If the deposit rate remains unchanged, the more rigorous capital requirements increase the weight on relatively expensive equity financing and raise banks' funding costs. But the deposit rates may be affected by the capital requirements. Banks would like to absorb more equity instead of deposits and the deposit rate decreases.

The old Basel frameworks introduce risk-weighted capital regulations which link capital to asset risk. But the risk-weighted capital regulations may exacerbate the procyclicality of lending (Danielsson et al. 2001; Kashyap and Stein 2004; Repullo and Suarez 2013). The rationale is that the capital requirements are enforced to a larger degree during an economy downturn. The banks constrained by the capital regulation are unable or unwilling to raise new equity in the crisis. They prefer to reduce risky loans and amplify the business cycle fluctuations. Then lending and credit availability are procyclical and amplify the business cycle (Gordy and Howells 2006; Freixas and Rochet 2008). The procyclical features of Basel II may contribute to the financial crisis (Brunnermeier 2009).

Behn et al. (2016) use the German banking data to examine how the model-based capital regulation affects the procyclicality of banking lending. They find that the countercyclicality of risk-weighted capital regulation has an important procyclical impact on banks' lending behaviors and the firms' access to funds. This chapter is the first empirical study to estimate how changes in risk-weighted capital regulation affect the supply of loans in the economy.

To solve this problem, the Basel III framework introduces a countercyclical capital buffer that reduces the pressure on banks to deleverage during the downturn. However, this may lead to higher risk in the banking system and cause unfavorable consequences for banks and economy.

Many economists attempt to identify the optimal capital requirement by quantifying the benefits and costs of capital requirements (Christiano and Ikeda 2016; Martinez-Miera and Suarez 2014; Corbae et al. 2019). Den Heuvel (2008) is the first to use a quantitative general equilibrium growth model to estimate the costs of higher capital requirements on welfare. The paper illustrates that higher capital requirements reduce financial fragility and lending amounts.

Since capital requirements have been imposed for about 30 years, a recent strand of literature empirically tests the effect of capital requirements on lending using different datasets. Aiyar et al. (2014) compare the credit distribution of the resident foreign branches with the one of the UK-owned banks which are subject to minimum capital requirements. Jiménez et al. (2017) exploit successive dynamic provisioning requirements imposed on Spanish banks in the 2000s. Fraisse et al. (2020) find that the capital requirements have a significant impact on lending. Facing the tough capital requirements, banks will shrink their risky assets and reduce lending.

95.4 Liquidity Regulation and Banks' Lending

The collapse of Lehman Brothers in September 2008 dragged other financial institutions into severe liquidity stress. Thus the need for routine liquidity management makes regulators realize that banks are exposed to liquidity risk. In addition, liquidity challenges that were experienced by the US and European banks rapidly spread into other markets through interconnections. However, banks face a dilemma on the right amount of liquidity to hold. The purpose of liquidity regulations is to ensure that banks hold adequate liquid assets in a crisis and finance their activities with stable funding sources.

A number of studies illustrate the importance of liquidity regulations. First, liquidity requirements are necessary because banks tend to hold insufficient liquidity because of the asymmetric information between depositors and banks (Diamond and Kashyap 2016). Banks have incentive to invest more in risky assets instead of liquid assets, while depositors can't get access to the perfect information about the banks' asset portfolios. As a result, banks hold insufficient liquidity and make themselves exposed to the bank-run risk. Liquidity regulations are required to restrict banks' behaviors. Second, liquidity regulations are justified firstly because of the market failures. The financial market cannot provide sufficient liquid assets when banks need more liquidity in a crisis. Liquidity requirements play an important role to avoid the liquidity squeeze. Third, liquidity regulations can supplement capital regulations in stabilizing the banking system (Calomiris et al. 2018). Finally, liquidity regulations would help to lower the chance of a bailout or a lender of last resort (LoLR) in a crisis. A bailout or an LoLR is costly because they induce moral hazard problems.

A main concern about liquidity regulations is that they are likely to harm banks' lending and profitability. The Basel III liquidity regulations are suspected to diminish banks' profitability by forcing banks to hold low-yield liquid assets and reduce lending (Banerjee and Mio 2018). Davis et al. (2020) develop a stylized interbank market environment and find that liquidity regulations reduce the incidence of bankruptcies but harm the investment efficiency. In the compound shock environment, liquidity regulations not only impose a loss of investment efficiency but also fail to reduce bankruptcies.

Although the importance of liquidity requirements had been discussed for a long time, the liquidity requirements hadn't been implemented until the launch of Basel III. As a result, the empirical studies about the effect of liquidity regulations are not as many as the capital-regulation ones. Now, there is still no conclusive empirical evidence about how liquidity regulations affect lending.

Bonner and Eijffinger (2016) study the Dutch banks and find that liquidity requirements push banks to borrow at longer terms and reduce their long-term lending. Because of the tighter interest margins by monetary policy, the Dutch banks do not change the lending rates and translate the funding costs to the lenders. Banerjee and Mio (2018) studies the impact on the UK banks and find that the limited interest rate impact is also appeared in the UK's banking system. The stricter

liquidity regulations affect banks' funding cost and have a negative impact on banks' lending and their profitability. Giordana and Schumacher (2017) investigates the impact of Basel III capital and liquidity requirements on default probability and profitability of banks in Luxembourg. They construct historical capital and liquidity ratios for the period from 2003 to 2011. They find that the impact of liquidity regulations on banks profitability is less clear-cut. The banks' funding structure is more important than the asset structure. Mashamba (2018) evaluates the effect of liquidity charges on the profitability of banks in emerging economies and find that the LCR has no detrimental effect on the performance of banks in emerging markets.

95.5 The Interaction Between Capital and Liquidity Regulations

The new Basel III intertwines the capital and liquidity regulations. The two kinds of regulations function on different parts of banks' balance sheets. Capital requirements directly restrict bank loans financed with deposits. In contrast, liquidity requirements require banks to maintain sufficient liquidity based on the deposit amounts. The two types of regulations drive banks to attract more equity and carry more liquid assets.

Liquidity and capital regulations work on bank asset and liability sides which are related to credit and liquidity risks (Bryant, 1980; Diamond and Dybvig, 1983). In the global financial crisis, credit and liquidity risks are demonstrated as the joint determinants of the financial intermediates' failure. Based on the interactive relationship between bank insolvency and illiquidity, Basel III imposes the joint regulations to protect banks and other intermediates in the future distress.

There is a growing literature on the interaction between liquidity and capital requirements. As most of theoretical models predict, the combination of equity capital and liquidity holding would stabilize the banking system. Vives (2014) uses a global game model to show that the regulations enhance financial stability by avoiding the bank run. Kara and Ozsoy (2021) argue that the liquidity regulation complements the capital requirement by allowing banks to hold more risky assets without raising fire sale risk. Walter (2019) argues that the joint regulations are efficient in internalizing the insolvency and liquidity risks.

But there are still arguments doubting the necessity of two regulations. Many economists prove that the combination of two requirements drive the implementation of Basel to be more costly. Liquidity requirements force banks to hold lower-yield liquid assets and reduce the profitable risky investments (Roger and Vlcek 2011). Capital requirements require banks to give up cheaper deposit funding and raise funding costs. The existence of both regulations reduce banks' revenues. Some banks points out that too much liquidity would reduce their solvency ability. The additional liquidity regulations harm banks' profits and raise the insolvency risk. If banks don't face high default risk, the negative effects of liquidity regulation will dominate.

Some economists prefer the liquidity requirements rather than the capital requirements. Adrian and Boyarchenko (2012) shows that liquidity requirements stabilize the financial system without harming the economic growth.

Others argue that capital requirements are sufficient in stabilizing the financial market. The new liquidity minimums in Basel III are imposed in the large banks that are already constrained by the capital requirements. The capital regulation has an effect on mitigating the liquidity risk. Some economists argue that liquidity and capital regulations are hard to be binding at the same time. Liquidity requirements tend to make capital less binding (Cecchetti and Kashyap 2018). As a result, the liquidity requirement may be redundant in stabilizing the financial market.

Many economists compare benefits and costs of two kinds of regulations. For example, Van Den Heuvel (2016) quantifies the welfare costs of requirements. According to this chapter, limited supply and important alternative uses of high-quality liquid assets make the liquidity requirement costly. But capital requirements carry higher social costs.

The interrelations of bank capital and liquidity requirements become complicated in empirically studies. For example, Schmaltz et al. (2014) simulate the likely response of banks to Basel III and find that banks will raise their stable deposit funding instead of investing in their liquid assets. However, Birn et al. (2017) show the opposite results. Distinguin et al. (2013) argues that the capital and liquidity are complements in large US banks. If large banks become illiquid, the regulatory capital ratios will decrease. But the regulatory capital ratios become higher if the smaller banks become illiquid. In the sample of small US banks, the capital and liquidity are substitutes. DeYoung et al. (2018) provide strong evidence that the small US bank can self-manage their liquidity without introducing the liquidity requirements. Their findings support the exemption of the smaller banks from the Basel III liquidity standards. According to their paper, imposing NSFR and LCR standards on small banks, like community banks, is redundant and unnecessarily costly. But the noncommunity banks may need the liquidity requirements in Basel III.

The noise in the empirical research is because that the liquidity requirements are just imposed in the financial markets. For example, the USA introduced LCR in 2015 and NSFR in 2016. Thus the impact of the joint requirements cannot be fully supported by data.

95.6 Conclusion

The Basel Accords are important for stabilizing the financial markets. Regulators should always discuss the present market conditions, learn lessons from the past, and update the Basel Accords. The global financial crisis prompts the introduction of capital and liquidity framework as main parts of Basel Accords. Following the policy development and implementation phases, it is now possible to assess the effects of the capital and liquidity regulations. This chapter reviews the recent academic evidence on this topic and summarizes major findings.

Overall, the literature to date concludes that banks respond to a binding capital or a liquidity constraint by raising long-term funding. Under normal economic conditions, banks have to bear more burdens. Their funding cost increases while risky

investments may be reduced. These effects have been illustrated by a range of theoretical models. But there is still insufficient empirical studies providing a clear conclusion about how the Basel Accords affect banks' behaviors.

Given that the LCR has been in effect in many countries since 2015, there is value in studying whether some of these model-implied effects are present in each country. As a starting point, one can measure the level of liquidity mismatch in the banking system, and document how it evolves with the LCR implementation following Bai et al. (2018), who construct a liquidity mismatch index (LMI) to gauge the difference between the market liquidity of assets and the funding liquidity of liabilities. One can also identify the extent to which capital and liquidity regulations would affect the loan amounts in each country and investigate the potential funding sources for small, mid-sized, and large banks to meet the standards.

It is meaningful to examine Basel III' effect on retail deposits. Recall that the LCR and NSFR treat retail deposits as a more stable funding source, which may be captured in their pricing. An interesting application would be to estimate this regulatory premium. In addition, how the new liquidity regulations work with other policies, like fiscal policy or monetary policy, is an interesting topic to explore. For example, Bech and Keister (2017) indicate that the introduction of an LCR requirement will make banks trade in interbank markets and affect interbank interest rates. Thus the regulations will change the impact of a central bank's monetary policy operations (OMO) which are used to control market interest rates. The model shows that steering market interest rates to a desired target level by open market operations may become more difficult when LCR requirement is implemented. The structure of an OMO becomes important since it may force banks to face LCR shortfall. The LCR requirement can change the impact of a central bank's OMO on the market interest rates. The central bank cannot use its pre-crisis operations to control market interest rates because of Basel III.

References

- Adrian, T., and N. Boyarchenko. 2012. Intermediary leverage cycles and financial stability. FRB of New York Staff Report
- Aiyar, S., C.W. Calomiris, and T. Wieladek. 2014. Does macro-prudential regulation leak? Evidence from a UK policy experiment. *Journal of Money, Credit and Banking* 46: 181–214
- Bai, J., A. Krishnamurthy, and C.H. Weymuller. 2018. Measuring liquidity mismatch in the banking sector. *Journal of Finance* 73: 51–93.
- Banerjee, R.N., and H. Mio. 2018. The impact of liquidity regulation on banks. *Journal of Financial Intermediation* 35: 30–44.
- Bech, M., and T. Keister. 2017. Liquidity regulation and the implementation of monetary policy. *Journal of Monetary Economics* 92: 64–77
- Begenau, J. 2020. Capital requirements, risk choice, and liquidity provision in a business-cycle model. *Journal of Financial Economics* 136: 355–378.

- Behn, M., R. Haselmann, and P. Wachtel. 2016. Procyclical capital regulation and lending. *Journal of Finance* 71: 919–956.
- Birn, M., M. Dietsch, and D. Durant. 2017. How to reach all Basel requirements at the same time? <https://ideas.repec.org/p/bfr/decfin/28.html>
- Bonner, C., and S.C. Eijffinger. 2016. The impact of liquidity regulation on bank inter-mediation. *Review of Finance* 20: 1945–1979.
- Brunnermeier, M.K. 2009. Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic Perspectives* 23: 77–100.
- Bryant, J. 1980. A model of reserves, bank runs, and deposit insurance. *Journal of Banking and Finance* 4: 335–344.
- Calomiris, C.W., F. Heider, and M. Hoerova. 2018. A theory of bank liquidity requirements. *SSRN Electronic Journal*.
- Cecchetti, S., and A. Kashyap. 2018. What binds? Interactions between bank capital and liquidity regulations. *The Changing Fortunes of Central Banking*: 192–202.
- Christiano, L., and D. Ikeda. 2016. Bank leverage and social welfare. *American Economic Review* 106: 560–564.
- Corbae, D., P. D'erasmo, and P. D. Erasmo. 2019. Capital requirements in a quantitative model of banking industry dynamics.
- Danielsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, H. S. Shin, et al. 2001. An academic response to Basel II.
- Davis, D.D., O. Korenok, J.P. Lightle, and E.S. Prescott. 2020. Liquidity requirements and the interbank loan market: An experimental investigation. *Journal of Monetary Economics* 115: 113–126.
- Deli, Y.D., and I. Hasan. 2017. Real effects of bank capital regulations: Global evidence. *Journal of Banking and Finance* 82: 217–228.
- Ven Den Heuvel, S.J. 2008. The welfare cost of bank capital requirements. *Journal of Monetary Economics* 55: 298–320.
- DeYoung, R., I. Distinguin, and A. Tarazi. 2018. The joint regulation of bank liquidity and bank capital. *Journal of Financial Intermediation* 34: 32–46.
- Diamond, D.W., and P.H. Dybvig. 1983. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*.
- Diamond, D.W., and A.K. Kashyap. 2016. Liquidity requirements, liquidity choice, and financial stability. In *Handbook of Macroeconomics* 2: 2263–2303.
- Diamond, D.W., and R.G. Rajan. 2001. Liquidity risk, liquidity creation, and financial fragility: A theory of banking. *Journal of Political Economy* 109: 287–327.
- Distinguin, I., C. Roulet, and A. Tarazi. 2013. Bank regulatory capital and liquidity: Evidence from US and European publicly traded banks. *Journal of Banking and Finance* 37: 3295–3317.
- Fraisse, H., M. Lé, and D. Thesmar. 2020. The real effects of bank capital requirements. *Management Science* 66: 5–23.
- Freixas, X., and J.-C. Rochet. 2008. Microeconomics of banking.
- Giordana, G.A., and I. Schumacher. 2017. An empirical study on the impact of Basel III standards on banks default risk: The case of Luxembourg. *Journal of Risk and Financial Management* 10: 8.
- Gordy, M.B., and B. Howells. 2006. Procyclicality in Basel II: Can we treat the disease without killing the patient? *Journal of Financial Intermediation* 15: 395–417.
- Hanson, S.G., A.K. Kashyap, and J.C. Stein. 2011. A macroprudential approach to financial regulation. *Journal of Economic Perspectives* 25: 3–28.
- Jiménez, G., S. Ongena, J.-L. Peydró, and J. Saurina. 2017. Macroprudential policy, countercyclical bank capital buffers, and credit supply: Evidence from the Spanish dynamic provisioning experiments. *Journal of Political Economy* 125: 2126–2177.
- Kara, G.I., and S.M. Ozsoy. 2021. Bank regulation under fire sale externalities. *Review of Financial Studies* 33: 2554–2584.

- Kashyap, A.K., and J.C. Stein. 2004. Cyclical implications of the Basel II capital standards. *Economic Perspectives* 28: 18–32.
- Kashyap, A., J. Stein, and S. Hanson. 2010. An analysis of the impact of substantially heightened capital requirements on large financial institutions. *Booth School of Business, University of Chicago, mimeo* 2: 1–47.
- Kelly, B., H. Lustig, and S. Van Nieuwerburgh. 2016. Too-systemic-to-fail: What option markets imply about sector-wide government guarantees. *American Economic Review* 106: 1278–1319.
- Kim, D., and A.M. Santomero. 1988. Risk in banking and capital regulation. *Journal of Finance* 43: 1219–1233.
- Koehn, M., and A.M. Santomero. 1980. Regulation of bank capital and portfolio risk. *Journal of Finance* 35: 1235–1244.
- Martinez-Miera, D., and J. Suarez. 2014. Banks endogenous systemic risk taking. *CEMFI Working Paper*.
- Mashamba, T. 2018. The effects of Basel III liquidity regulations on banks profitability. *Journal of Governance and Regulation* 7: 34–48.
- Modigliani, F., and M. Miller. 1963. *Corporate Income Taxes and the Cost of Capital: A Correction*. *American Economic Review* 53: 433–443.
- Repullo, R., and J. Suarez. 2013. The procyclical effects of bank capital regulation. *Review of Financial Studies* 26: 452–490.
- Roger, S., and J. Vlcek. 2011. Macroeconomic costs of higher Bank capital and liquidity requirements. *IMF Working Papers* 11: 1.
- Schmaltz, C., S. Pokutta, T. Heidorn, and S. Andrae. 2014. How to make regulators and shareholders happy under Basel III. *Journal of Banking and Finance* 46: 311–325.
- Van Den Heuvel, S. J. 2016. The Welfare effects of Bank liquidity and capital requirements.
- Vives, X. 2014. Strategic complementarity, fragility, and regulation.
- Walter, J. 2019. Bank capital regulation. *Global Financial Development Report 2019/2020* 105: 78–100.



Aron Gottesman

Abstract

This chapter introduces mortgage analysis. We first review mortgage characteristics and concepts. We then explore equations and Excel functions through which mortgage cash flows are calculated, including the mortgage payments, interest payments, and principal payments. This chapter then explains mortgage sensitivity and the use of Weighted Average Life (WAL) to measure mortgage sensitivity. This chapter concludes with an exploration of prepayment, including prepayment measurement and modelling using Single Month Mortality (SMM), Conditional Prepayment Rate (CPR), and Prepayment Speed Assumptions (PSA).

Keywords

Mortgages · Weighted-average life · Prepayment

JEL Classification

G21

96.1 Introduction

A mortgage is a lending agreement between a borrower and a lender where the borrower's real estate property is used as collateral. This chapter provides an introduction to mortgage analysis, concepts, and techniques. Among the questions that are addressed in this chapter are the following: What are the characteristics of a mortgage?

A. Gottesman (✉)

Department of Finance and Graduate Economics, Lubin School of Business, Pace University,
New York, NY, USA

e-mail: agottesman@pace.edu

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_98

2263

How are mortgage cash flows calculated? What Excel functions can be used to calculate mortgage cash flows? How is mortgage value sensitive to changes in mortgage rates? How is mortgage sensitivity measured? What is Weighted-Average Life (WAL) and what is its role? What is prepayment? How is prepayment measured, and modeled? What are Single Month Mortality (SMM), Conditional Prepayment Rate (CPR), and Prepayment Speed Assumptions (PSA), and what are their roles?

96.2 Mortgage Characteristics

Like a bond, in a mortgage the lender lends money to the borrower and the borrower pays interest and repays the money borrowed. The amount borrowed is referred to as the mortgage principal. The lender is referred to as the mortgagee and the borrower is referred to as the mortgagor. By a typical bond, the entire principal is paid at maturity. By a mortgage, the entire principal is not paid at maturity. Instead, through a process known as amortization, each payment that the mortgagor makes includes both principal and interest.¹

The length of the mortgage and the frequency of payments can vary. Common mortgage lengths are 15-year and 30-year mortgages.² Mortgages typically have monthly payments, though they occasionally have semi-monthly payments. The length of the mortgage is known as its “amortization term.” The amortization term is typically expressed in months, not years, hence a 15-year mortgage is referred to as a mortgage with a 180-month amortization term, and a 30-year mortgage is referred to as a mortgage with a 360-month amortization term.

The interest associated with a mortgage is known as its “mortgage rate” or “Annual Percentage Rate (APR).”³ Commonly, the frequency of compounding associated with the APR is equal to the frequency of the payments. Hence, the typical mortgage rate APR for a mortgage with monthly payments is compounded 12 times per year.

The mortgage rate may be a fixed rate of interest or a floating rate of interest that changes over time.⁴ When the mortgage rate is a floating rate of interest that changes over time, the mortgage is known as an “adjustable rate mortgage” or an ARM. The mortgage rate on an ARM is a function of an interest rate index. The period between

¹ Effectively, amortization represents an illiquid savings plan. Bernstein and Koudijs (2021) explore the important role that amortization has on household wealth building.

² Ye et al. (2014) study the relationship between mortgage rates and the choice of mortgage length.

³ McClatchey, C. and De la Torre, C. (2006) discuss other mortgage fees included in the APR and challenges associated with using APR when comparing mortgages.

⁴ Campbell (2013) Notes that long-term fixed-rate mortgages with minimal prepayment penalties hold an overwhelming majority of mortgage market share.

adjustments is the “adjustment period.” The date on which the mortgage rate adjusts is the “adjustment date.”⁵

ARMs often have initial rates that are low may have “adjustment caps” which are limits to the degree to which the mortgage rate can be adjusted on any given mortgage date, and may have “adjustment ceilings” which are limits to the level that the mortgage rate can reach, even over multiple adjustment periods. The ARM may allow the mortgagor to convert the mortgage into a fixed-rate mortgage. Such an ARM is known as a “convertible ARM.”⁶

When the mortgage rate is fixed, the mortgage rate is specified at the initiation of the mortgage and remains fixed during the amortization term. The mortgage payments made in each period consist of both interest and principal. When the mortgage rate is fixed, the total payment (i.e., the sum of the principal payment and interest payment) is the same amount every period.

An “amortization schedule” is a table that presents the mortgage payment and the interest and principal components of the mortgage payment for each month. Let’s explore the amortization schedule for the following example:

- Mortgage rate (APR_{12}): 6% (fixed rate, compounded 12 times per year)
- Frequency of mortgage payments: 12 per year (monthly)
- Amortization term: 180 months
- Mortgage principal: \$300,000

The amortization schedule for the first five and last five months of this mortgage is presented in Table 1.

As Table 1 shows, the mortgage payment (the second column of the amortization table) is \$2531.57 for each of the 180-months. In each month, the mortgagor is scheduled to make both interest and principal payments. Initially, the interest payments are a very large proportion of the mortgage payment. Towards the termination of the mortgage, the interest payments are a very small proportion of the mortgage payment.⁷

96.3 Calculating Mortgage Cash Flows

In the previous section, we learned that a mortgage payment represents the sum of the principal payment and interest payment associated with the given period. While the mortgage payment associated with a fixed-rate mortgage is the same every

⁵Bucks and Pence (2008) find that mortgage borrowers with adjustable-rate mortgages are likely to either underestimate their mortgage interest rate or be unaware of the amount by which their interest rate may change.

⁶Wang (2020) explores the role of ARM during the financial crisis of 2007–2009 and their decline in popularity.

⁷Calculation of the values specified in the amortization table will be explored in a subsequent section of this chapter.

Table 1 Amortization schedule for the first and last five months of a \$300,000 mortgage with a 180-month amortization term, monthly payments, and a fixed APR of 6% with monthly compounding

Month	Mortgage payment	Interest payment	Principal payment	Principal balance
0				\$300,000.00
1	\$2531.57	\$1500.00	\$1031.57	\$298,968.43
2	\$2531.57	\$1494.84	\$1036.73	\$297,931.70
3	\$2531.57	\$1489.66	\$1041.91	\$296,889.79
4	\$2531.57	\$1484.45	\$1047.12	\$295,842.67
5	\$2531.57	\$1479.21	\$1052.36	\$294,790.31
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
176	\$2531.57	\$62.35	\$2469.22	\$10,000.96
177	\$2531.57	\$50.00	\$2481.57	\$7519.39
178	\$2531.57	\$37.60	\$2493.97	\$5025.42
179	\$2531.57	\$25.13	\$2506.44	\$2518.98
180	\$2531.57	\$12.59	\$2518.98	\$0.00

period, the interest payment and principal payment differ each period. In this section, we will learn how to calculate the mortgage payment, the interest payment, the principal payment, and the principal balance at the end of the period.⁸

The present value of a series of cash flows can be calculated as follows:

$$PV = CF \cdot \left(\frac{1 - \frac{1}{(100\% + PER_m)^{n \cdot t_n}}}{PER_m} \right) \quad (1)$$

Where:

- PV = present value
- n = the number of times the cash flow is received
- CF = the cash flow that is received n times
- t_n = the number of years until the n th (final) cash flow takes place
- m = the frequency with which the cash flow is paid per year
- PER_m = the periodic effective rate that is compounded m times per year. The PER is the rate of interest for each compounding period.

We can use this expression to calculate the mortgage payment. To do so, recognize that when calculating a mortgage payment, our objective is *not* to identify the “present value.” We already know the present value – it’s the mortgage principal.

⁸While a typical mortgage borrower is unlikely to ever perform mortgage calculations in practice, it is interesting to note that Gerardi et al. (2013) find that numerical ability is negatively associated with mortgage default propensity.

Instead, when calculating a mortgage payment, our objective is to identify the cash flow. We can use the above equation to identify the mortgage payment as follows:

$$\text{Mortgage payment} = \text{Mortgage} / \left(\frac{1 - \frac{1}{(100\% + \text{PER}_m)^{m \cdot t_n}}}{\text{PER}_m} \right) \quad (2)$$

Unlike the mortgage payment, which is the same for every period, the interest payment is different for every period. The interest payment is calculated as follows:

$$\text{Interest payment} = \text{PER}_m \times \text{principal balance end of previous period} \quad (3)$$

Like the interest payment, the principal payment is different for every period. The principal payment is calculated as follows:

$$\text{Principal payment} = \text{Mortgage payment} - \text{Interest payment} \quad (4)$$

Since principal payments are made every period, the principal balance decreases with every period. The principal balance at the end of the period is calculated as follows:

$$\begin{aligned} &\text{Principal balance at the end of the period} \\ &= \text{Principal balance at the end of the previous period} - \text{Principal payment} \end{aligned} \quad (5)$$

For example, let's consider the cash flows associated with the second month of the mortgage presented in Table 1. The following are the inputs into the mortgage cash flow equations: $m = 12$, $\text{PER}_{12} = 6\%/12 = 0.5\%$, $t_n = 180/12 = 15$ years, and mortgage = \$300,000. The mortgage payment each month is $\$300,000 / \left(\frac{1 - \frac{1}{(100\% + 0.5\%)^{12 \cdot 15}}}{0.5\%} \right) = \$2,531.57$.

The second month's interest and principal payments are, respectively, $0.5\% \times \$298,968.43 = \$1,494.84$ and $\$2,531.57 - \$1,494.84 = \$1,036.73$. The remaining principal balance at the end of the second month is $\$298,968.43 - \$1,036.73 = \$297,931.70$.

Three functions in Excel through which mortgage cash flows can be calculated are the PMT, IPMT, and PPMT functions, which calculate the mortgage payment, interest payment, and principal payment, respectively.⁹ The inputs into these functions are PMT(rate, nper, and pv), IPMT(rate, per, nper, and pv), and PPMT(rate, per, nper, and pv), where:

- rate: The PER, compounded with the frequency that the mortgage payments take place per year
- nper: The number of payment periods over the amortization term

⁹Also see Arnold et al. (2017).

- **pv:** The mortgage principal, input with a negative sign.
- **per:** The period for which the interest or principal payment is being calculated.

Let's again consider the cash flows associated with the second month of the mortgage presented in Table 1. The following are the inputs into the Excel functions mortgage payment equation: $\text{rate} = 6\%/12 = 0.5\%$, $\text{nper} = 180$, $\text{pv} = \$300,000$, and $\text{per} = 2$. The mortgage payment, interest payment, and principal payment are calculated as follows: $\text{PMT}(0.5\%, 180, -300,000) = \2531.57 , $\text{IPMT}(0.5\%, 2, 180, -300,000) = \1494.84 , and $\text{PPMT}(0.5\%, 2, 180, -300,000) = \1036.73 .

96.4 Mortgage Sensitivity to Changes in Mortgage Rates

The value of a fixed-rate mortgage is the present value of the mortgage payments over the amortization term, discounted to the present using Eq. (1) at the current mortgage rate. The $\text{PV}()$ function in Excel can be used to discount a series of cash flows. In the context of mortgage valuation, the inputs are as follows: $\text{PV}(\text{rate}, \text{nper}, \text{and pmt})$, where rate and nper were defined earlier, and pmt is the mortgage payment, input with a negative sign.

Consider the mortgage presented in Table 1. In this case, the original mortgage rate was 6%, $\text{nper} = 180$, and $\text{pmt} = \$2531.57$. If the current mortgage rate is equal to the fixed mortgage rate of 6%, then the value of the mortgage presented in Table 1 is $\text{PV}(6\%/0.5\%, 180, -2531.57) = \$300,000$, or equal to the amount borrowed. If the current mortgage rate differs from the original mortgage rate, then the mortgage value will differ from the mortgage principal. To demonstrate, let's recalculate the mortgage value twice: assuming that the current mortgage rate is 5% and then again, assuming it is 7%. When 5%, the mortgage value is $\text{PV}(5\%/12, 180, -2531.57) = \$320,130.30$. At 7%, the mortgage value is $\text{PV}(7\%/12, 180, -2531.57) = \$281,652.24$.

The mortgage value's sensitivity to differences between the current mortgage rate and the original mortgage rate is similar to the sensitivity of a bond to difference between the yield and the bond's coupon rate. Specifically, if the current mortgage rate rises above the original mortgage rate, the value of the mortgage will decline. If the current mortgage rate falls below the original mortgage rate, the value of the mortgage will increase.

Further, the degree to which a mortgage will increase or decrease in value will depend on two key factors: The degree to which the current mortgage rate is different from the original mortgage rate and the amortization term of the mortgage. While any difference between the current mortgage rate and the original mortgage rate will result in a mortgage increasing or decreasing in value, the degree of increase or decrease depends on the degree to which the current mortgage rate is different from the original mortgage rate. If the current mortgage rate is similar to the original mortgage rate, the mortgage will not change in value very much. If the current mortgage rate is significantly different from the original mortgage rate, the mortgage will increase or decrease in value significantly. While any difference between the

current mortgage rate and the original mortgage rate will result in a mortgage increasing or decreasing in value, the degree of increase or decrease depends on the amortization term, with mortgages with longer amortization terms more sensitive.

Weighted-Average Life (WAL) is an index of mortgage sensitivity to changes in the mortgage rate. The higher that WAL is, the greater the sensitivity and vice versa. WAL is calculated as the principal-outstanding weighted average time until the mortgage payments associated with the mortgage are received.¹⁰ In equation form, WAL is specified as follows:

$$\text{WAL} = \sum_{i=1}^n t_i \cdot \frac{\text{Principal}_{t_i}}{\sum_{i=1}^n \text{Principal}_{t_i}} \quad (6)$$

where:

- n = the number of mortgage payment periods associated with the mortgage.
- t_i = the number of years until mortgage payment i occurs, where i ranges from 1 through n .
- Principal_{t_i} = the outstanding principal at the end of the payment period t_i , where i ranges from 1 through n .
- $\sum_{i=1}^n \text{Principal}_{t_i}$ = the sum of the outstanding principal at the end of all payment periods.

In this equation, $\frac{\text{Principal}_{t_i}}{\sum_{i=1}^n \text{Principal}_{t_i}}$ represents the “principal weight” of the principal outstanding at time t_i . The principal weights are multiplied by the years until each mortgage payment is received to determine the weighted average time until the mortgage payment associated with the mortgage is received.

Table 2 presents the calculation of the WAL for the mortgage presented in Table 2.

The first four columns of Table 2 are the amortization schedule. The sum of the fourth column, \$30,836,537.43 is the sum of the principal, $\sum_{i=1}^n \text{Principal}_{t_i}$. Each cell of the fifth column is principal weight for the given period, calculated as the principal outstanding at the end of the given period divided by the sum of the principal, $\frac{\text{Principal}_{t_i}}{\sum_{i=1}^n \text{Principal}_{t_i}}$. Each cell of the sixth column is the principal weight multiplied by the

¹⁰ Also see Yesayan (2009).

Table 2 Weighted-Average Life (WAL) calculations for a \$300,000 mortgage with a 180-month amortization term, monthly payments, and a fixed APR of 6% with monthly compounding; first and last five months are shown

Month	Mortgage payment	Interest payment	Principal payment	Principal balance	Principal weight	Principal weight x years
0				\$300,000.00		
1	\$2531.57	\$1500.00	\$1031.57	\$298,968.43	0.970%	0.0008
2	\$2531.57	\$1494.84	\$1036.73	\$297,931.70	0.966%	0.0016
3	\$2531.57	\$1489.66	\$1041.91	\$296,889.79	0.963%	0.0024
4	\$2531.57	\$1484.45	\$1047.12	\$295,842.67	0.959%	0.0032
5	\$2531.57	\$1479.21	\$1052.36	\$294,790.31	0.956%	0.0040
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
176	\$2531.57	\$62.35	\$2469.22	\$10,000.96	0.032%	0.0048
177	\$2531.57	\$50.00	\$2481.57	\$7519.39	0.024%	0.0036
178	\$2531.57	\$37.60	\$2493.97	\$5025.42	0.016%	0.0024
179	\$2531.57	\$25.13	\$2506.44	\$2518.98	0.008%	0.0012
180	\$2531.57	\$12.59	\$2518.98	\$0.00	0.000%	0.0000
			Sum of principal =	\$30,836,537.43	WAL =	5.3762

years until the given mortgage payment, $t_i \cdot \frac{\text{Principal}_{t_i}}{\sum_{i=1}^n \text{Principal}_{t_i}}$. The sum of the sixth column,

5.3762, is the $\text{WAL} = \sum_{i=1}^n t_i \cdot \frac{\text{Principal}_{t_i}}{\sum_{i=1}^n \text{Principal}_{t_i}}$.

96.5 Prepayment

A characteristic of residential mortgages is that the mortgagee has the right to prepayment, that is, to repay mortgage principal each month, above and beyond the principal payment scheduled for the given month. The mortgagee can choose to prepay any amount of principal. For example, if the current mortgage rate falls below the given mortgage’s original mortgage rate, the mortgagee may prepay the entire mortgage. Alternatively, the mortgagee may choose to prepay a small amount of principal each month.

Single Month Mortality (SMM) is a measure of prepayment in each month. It is calculated as follows:

$$\text{SMM} = \frac{\text{Prepayment in given month}}{\text{Principal balance following the scheduled principal payment in given month}}$$

(7)

The SMMs for months 1, 2, and 3 of the mortgage presented in Table 1 are calculated, respectively, as $\frac{\$1,000}{\$298,968.43} = 0.3345\%$, $\frac{\$1,000}{\$296,935.17} = 0.3368\%$, and $\frac{\$1,000}{\$294,900.24} = 0.3391\%$

The Conditional Prepayment Rate (CPR) is an annualized measure of prepayment of the prepayment rate. It is calculated as follows:

$$\text{CPR} = 1 - (1 - \text{SMM})^{12} \quad (8)$$

If the CPR is known, we can invert this equation and calculate the SMM as follows:

$$\text{SMM} = 1 - (1 - \text{CPR})^{1/12} \quad (9)$$

The CPRs for months 1, 2, and 3 of the mortgage presented in Table 1 are calculated, respectively, as $1 - (1 - 0.3345\%)^{12} = 3.9408\%$, $1 - (1 - 0.3368\%)^{12} = 3.9673\%$, and $1 - (1 - 0.3391\%)^{12} = 3.9941\%$

The Standard Prepayment Model of The Bond Market Association makes certain “Prepayment Speed Assumptions,” known as the “PSA100.” The PSA100 assumptions are as follows: In the first month, the CPR is 0.2%. The CPR then increases by 0.2% each month until the 30th month. The CPR remains 6% until the mortgage is completely paid off.

PSA100 is perceived as a “benchmark” of prepayment. In reality, the actual prepayment may take place quicker or slower. If prepayment is quicker than the assumed prepayment in the PSA100, the “100” in the PSA100 name is replaced with a number greater than 100. Conversely, if prepayment is slower than the assumed prepayment in the PSA100, the “100” in the PSA100 name is replaced with a number less than 100.

For example, PSA200 refers to an assumed prepayment schedule that is twice as quick as the assumed prepayment in the PSA100. Hence, in the first month, the CPR was 0.4%. The CPR then increases by 0.4% each month until the 30th month. The CPR remains 12% until the mortgage is completely paid off.

Tables 3 and 4 present the mortgage provided in Table 1, assuming PSA100 and PSA 700. Observe that the WAL in the PSA100 scenario (Table 3) is 4.6467, while the WAL in the PSA700 scenario (Table 4) is 1.9663.

96.6 Conclusions

This chapter introduced mortgage analysis. The chapter first reviewed mortgage characteristics and concepts, including the concepts of mortgage principal, mortgagee, mortgagor, amortization, amortization term, payment frequency, mortgage rate, compounding, and amortization schedules. The chapter also described adjustable rate mortgages (ARM) and fixed-rate mortgages.

This chapter next explored equations through which mortgage cash flows are calculated, including the mortgage payments, interest payments, principal payments,

Table 3 PSA100 amortization schedule. Calculations are for a \$300,000 mortgage with a 180-month amortization term, monthly payments, and a fixed APR of 6% with monthly compounding; first and last five months are shown

Month	Mortgage payment	Interest payment	Scheduled principal payment	Principal balance following the scheduled principal payment	CPR	SMM	Additional principal prepayment	Principal balance following the prepaid principal payments	Principal weight	Principal weight x years
0				\$300,000.00						
1	\$2531.57	\$1500.00	\$1031.57	\$298,968.43	0.20%	0.0167%	\$49.87	\$298,918.56	1.234%	0.0010
2	\$2531.15	\$1494.59	\$1036.56	\$297,882.00	0.40%	0.0334%	\$99.48	\$297,782.52	1.230%	0.0020
3	\$2530.30	\$1488.91	\$1041.39	\$296,741.13	0.60%	0.0501%	\$148.78	\$296,592.35	1.225%	0.0031
4	\$2529.03	\$1482.96	\$1046.07	\$295,546.28	0.80%	0.0669%	\$197.76	\$295,348.52	1.220%	0.0041
5	\$2527.34	\$1476.74	\$1050.60	\$294,297.92	1.00%	0.0837%	\$246.38	\$294,051.54	1.214%	0.0051
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
176	\$1107.44	\$27.28	\$1080.17	\$4374.95	6.00%	0.5143%	\$22.50	\$4352.45	0.018%	0.0026
177	\$1101.75	\$21.76	\$1079.98	\$3272.46	6.00%	0.5143%	\$16.83	\$3255.63	0.013%	0.0020
178	\$1096.08	\$16.28	\$1079.80	\$2175.83	6.00%	0.5143%	\$11.19	\$2164.64	0.009%	0.0013
179	\$1090.44	\$10.82	\$1079.62	\$1085.02	6.00%	0.5143%	\$5.58	\$1079.44	0.004%	0.0007
180	\$1084.83	\$5.40	\$1079.44	\$0.00	6.00%	0.5143%	\$0.00	\$0.00	0.000%	0.0000
							Sum of principal =	\$ 24,218,411.49	WAL =	4.6467

Table 4 PSA700 amortization schedule. Calculations are for a \$300,000 mortgage with a 180-month amortization term, monthly payments, and a fixed APR of 6% with monthly compounding; first and last five months are shown

Month	Mortgage payment	Interest payment	Scheduled principal payment	Principal balance following the scheduled principal payment	CPR	SMM	Additional principal prepayment	Principal balance following the scheduled and prepaid principal payments	Principal weight	Principal weight x years
0				\$300,000.00						
1	\$2531.57	\$1500.00	\$1031.57	\$298,968.43	1.40%	0.1174%	\$351.05	\$298,617.37	3.086%	0.0026
2	\$2528.60	\$1493.09	\$1035.51	\$297,581.86	2.80%	0.2364%	\$703.43	\$296,878.43	3.068%	0.0051
3	\$2522.62	\$1484.39	\$1038.23	\$295,840.20	4.20%	0.3569%	\$1055.92	\$294,784.28	3.047%	0.0076
4	\$2513.62	\$1473.92	\$1039.70	\$293,744.58	5.60%	0.4791%	\$1407.30	\$292,337.28	3.021%	0.0101
5	\$2501.57	\$1461.69	\$1039.89	\$291,297.39	7.00%	0.6029%	\$1756.32	\$289,541.07	2.992%	0.0125
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
176	\$1.84	\$0.05	\$1.80	\$7.29	42.00%	4.4379%	\$0.32	\$6.96	0.000%	0.0000
177	\$1.76	\$0.03	\$1.73	\$5.24	42.00%	4.4379%	\$0.23	\$5.00	0.000%	0.0000
178	\$1.68	\$0.03	\$1.66	\$3.34	42.00%	4.4379%	\$0.15	\$3.20	0.000%	0.0000
179	\$1.61	\$0.02	\$1.59	\$1.60	42.00%	4.4379%	\$0.07	\$1.53	0.000%	0.0000
180	\$1.54	\$0.01	\$1.53	\$0.00	42.00%	4.4379%	\$0.00	\$0.00	0.000%	0.0000
							Sum of principal =	\$ 9,675,709.27	WAL =	1.9663

and remaining principal balance. This chapter also explored and the Excel functions PMT(), IPMT(), and PPMT() through which these mortgage cash flows can be calculated.

This chapter then explains mortgage sensitivity. It demonstrated that if the current mortgage rate rises above the original mortgage rate, the value of the mortgage will decline. If the current mortgage rate falls below the original mortgage rate, the value of the mortgage will increase. The use of Weighted-Average Life (WAL) to measure mortgage sensitivity was also explored. It was shown that WAL is calculated as the principal-outstanding weighted average time until the mortgage payments associated with the mortgage are received.

This chapter concludes with an exploration of prepayment, which is the mortgagee's right to prepayment. It was shown that Single Month Mortality (SMM) is a measure of prepayment in a given month; the Conditional Prepayment Rate (CPR) is an annualized measure of prepayment of the prepayment rate; and that the Standard Prepayment Model of The Bond Market Association makes certain "Prepayment Speed Assumptions," known as the "PSA100."

References

- Arnold, T., J.H. Earl, and C.D. Marshall. 2017. Using Google sheets to determine mortgage information. *The Journal of Wealth Management* 19 (4): 128–131.
- Bernstein, A., and P. Koudijs 2021. *The mortgage Piggy Bank: Building wealth through amortization*. NBER Working Paper No. 28574.
- Bucks, B., and K. Pence. 2008. Do borrowers know their mortgage terms? *Journal of Urban Economics* 64 (2): 218–233.
- Campbell, J.Y. 2013. Mortgage market design. *Review of Finance* 17: 1–33.
- Gerardi, K., L. Goette, and S. Meier. 2013. Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences* 110: 11267–11271.
- McClatchey, C., and C. De la Torre. 2006. Comparing fixed-rate mortgage loans via the annual percentage rate: Cautions and caveats. *Journal of Financial Service Professionals* 60: 16–18.
- Wang, R.P. 2020. Cause determination of the adjustable-rate mortgage market collapse during the financial crisis. *International Journal of Economics and Finance* 12 (11): 83–90.
- Ye, G., G. Deng, and Z. Li. 2014. Mortgage rate and choice of mortgage length: A quasi-experimental evidence from Chinese transaction-level data. *Journal of Housing Economics* 25: 96–103.
- Yesayan, A. 2009. *Mortgage pricing*. SSRN working paper. Available at <https://ssrn.com/abstract=1414351>



A History of Commercially Available Risk Models

97

John Blin, John Guerard, and Andrew Mark

Abstract

Multi-factor risk models have been used in portfolio selection since the 1960s and early 1970s. The work of Barr Rosenberg led to the creation of Barra, the first major commercially available portfolio selection software. Rudd and Clasing brought Barra to the academic audience as professional managers were embracing it in industry. It is important to see how quantitative analysis was developed in investment research and analysis. Harry Markowitz, Bill Sharpe, Jan Mossin, and Jack Treynor pioneered capital market equilibrium and the creation and estimation of the Capital Asset Pricing Model. In the 1970s, multi-factor risk models were developed and estimated by Barr Rosenberg, Andrew Rudd, and their colleagues at Barra; John Blin and Steve Bender at APT; and Sebastian Ceria at Axioma. Multi-factor models pushed out the efficient frontier relative to a single factor risk model. McKinley Capital Management developed its Horse Race risk model philosophy to assess the most effective risk models for maximizing the Geometric Means, Sharpe Ratios, and Information Ratios. Several special issues of practitioner-oriented applied investment research featured the McKinley Horse Race results and reported that properly constructed multi-factor risk models created portfolios in which tilt variables that have statistically significant ICs produced statistically significant Active Returns. One must remove the market effect and extra-market covariances to properly estimate the contribution of a variable to the creation of efficient portfolios. APT and Axioma have reigned

J. Blin

APT Investment Management, Hong Kong, China

J. Guerard (✉)

McKinley Capital Management, LLC, Computational Finance and Risk Management Program,
The University of Washington, Anchorage, AK, USA

e-mail: jguerard@mckinleycapital.com

A. Mark

GlobeFlex, LP, San Diego, CA, USA

supreme in those tests. The integration of Axioma and AAF and ITG transactions costs on the FactSet Investment have led MCM to implement an Axioma portfolio.

Keywords

Optimized Portfolios · Risk Models · Earnings Forecasting

97.1 Introduction

Investing in stocks has always been risky, and the Crash of 1929 still has a solid place in US history. Yet over the last 70 years, the work of mathematicians has made it possible for managers to deal with that risk. In this chapter, we take a look at how this unfolded, and then we focus on three of today's leading risk-management systems.

In a breakthrough article published in 1952 in the *Journal of Finance*, Harry Markowitz introduced the portfolio risk equation. At heart, it was a simple quadratic constrained optimization problem. By then, linear programming was well understood – i.e., how to deal with the constrained optimization of linear systems. Likewise, quadratic extension methods had also been studied. Markowitz's paper was truly novel in that it applied its findings to financial portfolios. A clear-cut example of applied mathematics at its most practical.

As Markowitz recounts, when he was looking for a thesis topic, a chance conversation with a stockbroker got him thinking about the following: How do you take a group of recommended stocks and create a “good” portfolio? Was such a portfolio simply a collection of “good” stocks? Everyone knew in some form or other that “a wise man does not venture all his eggs in one basket.” But Don Quixote never said what to do about the eggs and basket. In his paper, Markowitz did more than hint about what to do, and eventually a new industry was born. Today a number of commercial risk management systems are available. Here we will be looking at three of them: Barra, APT, and Axioma.

Let's start with a quick review of the academic efforts. With his *Portfolio Selection: Efficient Diversification of Investments* (1959), Markowitz zeroed in on the risk-return trade-off. He's been expanding this effort ever since, including in his *Risk-Return Analysis* series (2013, 2016, and 2019, with a forthcoming fourth volume). Here we refer to the Markowitz Mean-Variance Model as the “MMV.” The Sharpe Diagonal Model drastically reduced the computational requirements of the MMV. It was followed by the Sharpe-Lintner-Mossin (1964, 1965, 1966) forms of the Capital Asset Pricing Model (CAPM). Estimation issues for CAPM were reported in Black et al. (1972). Aside from asset pricing models, multi-factor risk models also sprang up: King (1966), Cohen and Pogue (1967), Farrell (1974, 1997), Stone (1974), Rosenberg and McKibben (1973), Rosenberg (1974), Rosenberg and Marathe (1979), Ross (1976), Roll and Ross (1984), Rudd and Clasing (1982), Dhrymes et al. (1984, 1985), and Fama and French (1992, 1995, and 2008).

Commercial implementations soon followed: First, Barra in 1979. Then, APT (Advanced Portfolio Technologies, Inc.) in 1987. And finally, Axioma, created in the late 1990s and gaining practitioner acceptance in the 2000–2019 time period. Bloomberg will join the fray when it releases its own risk model.

In what follows we introduce readers to the three practitioners' multi-factor models and their creators: Specifically, Barra developed fundamentally based domestic and global risk models that are estimated from prespecified fundamental variables. APT developed multi-factor domestic and global risk models using principal components analysis (PCA). PCA models extract statistical factors (principal components) from the returns data, while fundamentally based models start with prespecified fundamental variables. Axioma created both fundamentally and statistically based risk models. Recently, Bloomberg has been preparing to release its own fundamentally based risk model.

We end by discussing the McKinley Capital Management (MCM) Horse Races for the 2010–2019 time period. These races were used to evaluate stock-selection models driven by (1) earnings forecasting and (2) composite variables of value, price momentum, and earnings forecasting. The portfolios built with those two sets of variables were optimized for minimum risk using the three risk models highlighted in this chapter. Thus, they are a joint test of the power of such variables *as long as* the portfolio risk is tightly controlled by different risk models.

97.2 Risk and Returns of Portfolios

In Chaps. 7 and 8 of *Portfolio Selection* (1959), Markowitz shows how to build portfolios that minimize risk. With such portfolios, no stock reshuffling can further reduce the overall portfolio risk. Mathematically, to minimize risk, each stock's first partial derivative (change in portfolio risk as the weight varies) must equal zero. In short, no increase or decrease in any stock weight can further reduce the risk of the portfolio. While varying those weights generates an infinity of feasible portfolios, only a subset of those is "efficient" – i.e., maximizing return for a given risk level. That risk is measured by the portfolio's standard deviation. The portfolio's expected return, $E(R_p)$, is the weighted sum of the stocks' expected returns, the weights being the portfolio weights. The portfolio's standard deviation is the weighted sum of the variances and covariances:

$$E(R_p) = \sum_{i=1}^N x_i E(R_i) = \sum_{i=1}^N x_i \mu_i \quad (1)$$

$$\sigma_p^2 = \sum_{i=1}^N \sum_{j=1}^N x_i x_j c_{ij} \quad (2)$$

where μ is the expected return vector, C is the variance-covariance matrix, and x is the vector of portfolio weights. The efficient frontier is traced out by the solutions to the problem:

$$\text{minimize}_{\{x_i \geq 0, x_i \leq \bar{u}\}} x^T Cx - \lambda \mu^T x \quad (3)$$

as λ varies to reflect the risk-return trade-off parameter and $\bar{u}$ is the fixed upper bound.

In 1970, Sharpe introduced the Capital Market Line (CML) to describe the return-risk trade-off when investors can borrow and lend at the risk-free rate and are compensated for bearing risk. Further, investors have a one-period horizon, pay no taxes or transactions costs (in their dreams), share the same views about stocks, and in all cases seek to hold mean-variance efficient portfolios. Specifically, all investors have identical probabilities for the distribution of future stock returns:

$$E(\tilde{R}_P) = R_F + \left[\frac{E(\tilde{R}_M) - R_F}{\sigma_M} \right] \sigma_P \quad (4)$$

where:

$E(R_P)$ = expected return on the portfolio

$E(R_M)$ = expected return on the market portfolio, where all securities are held relative to their market value

R_F = risk-free rate

σ_M = standard deviation of the market portfolio

σ_P = standard deviation of the portfolio

Note that as the portfolio's standard deviation rises, its expected return must rise.

97.3 An Introduction to Modern Portfolio Theory: The Capital Asset Pricing Model (CAPM)

Markowitz built a portfolio construction theory in which investors should be compensated with higher returns for bearing higher risk. The Markowitz framework measured risk as the portfolio's standard deviation – that is, its total risk. Further analysis of CAPM by Sharpe (1964), Lintner (1965), and Mossin (1966), as well as by Jack Treynor, shows that investors are not compensated for assuming total risk, but only for assuming *market* risk (*systematic* risk) as measured by the stock's beta. As a reminder, the beta is the slope of the market model in which the stock return is regressed on the market return. If risk can be diversified away in an optimized portfolio, investors will not be compensated for that risk. They will be compensated only for the portfolio's systematic risk.

CAPM holds that a stock's return is a function of its beta:

$$R_{jt} = R_F + \beta_j [E(R_{Mt}) - R_F] + e_j \quad (5)$$

where:

R_{jt} = expected security return at time t

$E(R_{Mt})$ = expected return on the market at time t

R_F = risk-free rate

β_j = security beta

e_j = randomly distributed error term

Now let's focus on the CAPM beta. It's the model's measure of systematic risk and it uses the Capital Market Line equilibrium condition:

$$\beta_j = \frac{\text{Cov}(R_j, R_M)}{\text{Var}(R_M)} \quad (6)$$

$$\begin{aligned} E(R_j) &= R_F + \left[\frac{E(R_M) - R_F}{\sigma_M^2} \right] \text{Cov}(R_j, R_M) \\ &= R_F + [E(R_M) - R_F] \frac{\text{Cov}(R_j, R_M)}{\text{Var}(R_M)} \\ E(R_j) &= R_F + [E(R_M) - R_F] \beta_j \end{aligned} \quad (7)$$

The Security Market Line, SML, shown in Eq. (7), captures the linear relationship between return and systematic risk as measured by beta. The second pass regression, regressing future returns on the estimated beta coefficients, produced the statistical evidence that led Black et al. (1972) to question CAPM's empirical validity.

97.4 The Barra Risk Model

As today's oldest commercially available risk model, the Barra model dates back to the mid-1970s, with many subsequent updates as described below. Note that Barra Inc. was acquired in 2004 by Morgan Stanley Capital International Inc.¹

Following publication of Black et al. (1972), multi-factor risk models became a hot topic among academics and practitioners, and the risk model developed by Barra was among the first to be commercially available. It was based on the academic research published in Barr Rosenberg and Walt McKibben (1973), Rosenberg (1974), Rudd and Rosenberg (1980), Rosenberg and Marathe (1979), and Rudd and Clasing (1982). Barr Rosenberg marketed the Barra risk model in academic circles and presented it at institutional investor meetings. As a result, the Barra risk model became the primary institutional risk model of the 1975–2005 period.

¹ Rudd and Rosenberg (1979) published one of the first academic/commercial portfolio optimization analyses in the Elton and Gruber monograph to honor Harry Markowitz. Miller, Xu, and Guerard (2014) used the Barra optimizer.

Accurate characterization of portfolio risk requires an accurate estimate of the covariance matrix of security returns. A relatively simple way to estimate this covariance matrix is to use the history of security returns to compute each variance, covariance, and security beta. Using beta (the covariance of security and market index returns) can produce an estimate of a corporation's cost of equity. However, this simple approximation doesn't yield a very accurate estimate. The simple, single-index beta estimation approach suffers from two major drawbacks:

- Estimating a covariance matrix for the Russell 3,000 stocks requires a great deal of data.
- The approach is subject to estimation error. For example, in Guerard and Schwartz (2007), the reader saw that the estimated correlation between two stocks such as DuPont and IBM had a higher correlation than that between DuPont and Dow. This is counterintuitive: One would normally expect a higher correlation between DuPont and Dow than between DuPont and IBM given that DuPont and Dow were both chemical firms.²

Taking this further, one can argue that companies with similar characteristics, such as those in the same line of business, should have returns that behave similarly. For example, DuPont, IBM, and Dow Chemical will all have a common component in their returns because of their sensitivity to news that affects the stock market, as measured by their respective betas.

By beginning with this intuition about the sources of co-movement in stocks' returns, Rosenberg (1974) made substantial progress in estimating the covariance matrix of stock returns. Specifically, it estimates three components: the covariance matrix of common sources of stock returns, the variances of stock-specific returns, and the sensitivity of stock returns to the common sources of variation in their returns. Together they form the Barra risk model. Because the common sources of risk are likely to be much fewer than the number of stocks, we can get by with a much smaller covariance matrix. This means we require a much smaller history of returns. Moreover, because similar stocks are going to have higher sensitivities to similar common sources of risk, similar stocks will be more highly correlated than dissimilar stocks.

The Barra risk model is a multiple-factor model (MFM). MFMs build on single-factor models by including and describing the interrelationships among factors.³

For single-factor models, the equation that describes the excess rate of return is:

$$\tilde{r}_j = X_j^f + \tilde{u}_j \quad (8)$$

²Dow and DuPont merged in 2015, whereas the correlation comments were made in 2007.

³The reader is referred to the BARRA (US-E3) United States Equity, Version 3, Risk Model Handbook. Guerard and Mark (2003 and 2020) and Miller, Xu, and Guerard (2014) used the US-E3 model in their analysis. The reader is also referred to Rosenberg and Marathe (1979), Rudd and Rosenberg (1980), Rudd and Clasing (1982), and Grinold and Kahn (1999).

where:

- $\tilde{r}_j$ = total excess return over the risk-free rate
- X_j = sensitivity of stock j to the factor
- $\tilde{f}_j$ = rate of return on the factor
- $\tilde{u}_j$ = nonfactor (specific) return on stock j

We can expand this model to include K factors. The total excess return equation for a multiple-factor model becomes:

$$\tilde{r}_j = \sum_{k=1}^K X_{jk} \tilde{f}_k + \tilde{u}_j \quad (9)$$

where:

- X_{jk} = risk exposure of stock j to factor k
- $\tilde{f}_k$ = rate of return on factor k

Note that when $K = 1$, the MFM equation reduces to the earlier single-factor version – the CAPM discussed in the previous chapter.

When a portfolio consists of only one stock, Eq. (3) describes its excess return. But most portfolios comprise many stocks, each representing a proportion, or weight, of the total portfolio. When weights $h_{p1}, h_{p2}, \dots, h_{pN}$ reflect the proportions of N stocks in portfolio P, we express the excess return in the following equation:

$$\tilde{r}_p = \sum_{k=1}^K X_{Pk} \tilde{f}_k + \sum_{j=1}^N h_{pj} \tilde{u}_j \quad (10)$$

where:

$$X_{Pk} = \sum_{j=1}^N h_{pj} X_{jk}$$

This equation includes the risk from all sources and lays the groundwork for further MFM analysis.

Investors look at the variance of their total portfolios to make a comprehensive assessment of risk. To calculate the variance of a portfolio, one needs to calculate the covariances of all the constituent components. Without the framework of a multiple-factor model, estimating the covariance of each asset with every other asset would be a computational nightmare and subject to significant estimation errors. With this in mind, let's examine the risk structure of the Barra MFM.

The Rosenberg MFM framework was developed and estimated in Rosenberg's study with McKibben (1973) and the Rosenberg extra-market component study (1974), in both of which stock-specific risk could be modeled as a function of financial descriptors or financial characteristics of the firm. The financial characteristics that were statistically associated with beta during the 1954–1970 period were:

1. Latest annual proportional change in earnings per share
2. Liquidity, as measured by the quick ratio
3. Leverage, as measured by the senior-debt-to-total-assets ratio
4. Growth measure of earnings per share
5. Book-to-price ratio
6. Historical beta
7. Logarithm of stock price
8. Standard deviation of earnings per share growth
9. Gross plant per dollar of total assets
10. Share turnover

The statistically significant determinants of the security systematic risk became the basis of the Barra E1 Model risk indices. Rudd and Clasing (1982) provides a complete discussion of the initial US-E1 Barra system.

By 1998, the domestic Barra E3 (US-E3) model had grown to include 13 sources of factor, or systematic, exposures: volatility, momentum, size, size nonlinearity, trading activity, growth, earnings yield, value, earnings variation, leverage, currency sensitivity, dividend yield, and non-estimation universe. The multiple-factor model significantly reduces the number of calculations. For example, in the US Equity Model (US-E3), 65 factors capture the risk characteristics of equities. This reduces the number of covariance and variance calculations; moreover, since there are fewer parameters to determine, they can be estimated with greater precision. The Barra US-E3 factor index definitions are shown in Appendix 1. The Barra risk management system begins with the MFM equation:

$$\tilde{r} = X\tilde{f} + \tilde{u} \quad (11)$$

where:

$\tilde{r}$ = excess return on asset I

X = exposure coefficient on the factor

$\tilde{f}$ = factor return

$\tilde{u}$ = specific return

Substituting this relationship in the basic equation, we find that:

$$\text{Risk} = \text{Var}(\tilde{r}) \quad (12)$$

$$= \text{Var}(X\tilde{f} + \tilde{u}) \quad (13)$$

Using the matrix algebra formula for variance, the risk equation becomes:

$$\text{Risk} = XFX^T + \Delta \quad (14)$$

where:

X = exposure matrix of companies to factors

F = covariance matrix of factors

X^T = transpose of X matrix

Δ = diagonal matrix of specific risk variances

This is the basic equation that defines the matrix calculations used in risk analysis in the Barra equity models.

Let's look at three examples of applying Barra risk models to portfolio construction and management. We begin with the results of the Guerard and Mark (2003) study, which used estimated earnings forecasting components of a composite model (CTEF) for the Russell 3000 universe during the 1990–2001 period. This composite model combined four earnings forecasting yields: the 1-year-ahead and 2-year-ahead forecasted earnings per share divided by the stock price, the monthly 1-year-ahead and 2-year-ahead earnings revisions, the monthly 1-year-ahead and 2-year-ahead breadth, and the number of analysts raising EPS estimates *less* the number of analysts lowering EPS estimates *divided by* the total number of analyst estimates.⁴ The CTEF model produced higher information coefficients (ICs) as well as higher and more statistically significant asset selections than its components in the Russell 3000 universe.⁵ Guerard and Mark (2003) and Guerard and Schwartz (2007) reported the effective performance of domestic and international CTEF portfolios. Table 1 shows the earnings component results for the Russell 3000 for portfolios of about 100 stocks, each tilted on individual and component CTEF factors.

The CTEF variable produces statistically significant total active returns and asset selection. The CTEF variable loading on the risk index is statistically significant at the 10% level because of its loading on the earnings yield and nonlinear size indexes, as was the case with its breadth components. There are no statistically significant sector exposures in the CTEF variable. The CTEF model offers statistically significant asset selection in a multiple-factor model framework.

⁴The reader is referred to Bruce and Epstein (1994) for a collection of published earnings forecasting research during the 1968–1992 time period, and to Brown (1998 and 2008) for a comprehensive annotated bibliography of this research.

⁵The forecast earnings per share for the 1-year-ahead and 2-year-ahead periods, FEP1 and FEP2, offer negative, but statistically insignificant, asset selection. The total active returns are positive, and not statistically significant. The asset selection is negative, because the FEP variables have positive and statistically significant loadings on the risk indices – particularly the earnings yield index. The factor loading of the FEP variables on the earnings yield risk index is not unexpected given that the earnings yield factor index in the US-E3 includes the forecast earnings-to-price variable. Thus, there is no multiple-factor model benefit to the FEP variables. The monthly revision variables, the RV variables, offer no statistically significant total active returns, or asset selection abilities. The breadth variables, BR, produce statistically significant total active returns and asset selection, despite statistically significant risk index loadings. The breadth variable loads on the earnings yield and growth risk indices. Taking a closer look at the BR1 factor risk index loading, the BR1 variable leads a portfolio manager to have a positive average active exposure to the earnings yield index, which incorporates the analyst predicted earnings-to-price and historical earnings-to-price measures. The BR1 tilt has a negative and statistically significant average exposure to size, nonlinearity, and the cube of normalized market capitalization. This result is consistent with analyst revisions being more effective in smaller-capitalized securities.

Table 1 Components of the composite earnings forecasting variable, 1990–2001

R3000 earnings analysis	Total active	T-stat	Asset Selection	T-stat	Risk Index	T-stat
FEP1	2.14	1.61	−1.18	−1.17	4.20	4.42
FEP2	1.21	0.91	−1.43	−1.35	3.33	3.35
RV1	0.76	0.69	0.34	0.42	0.92	1.46
RV2	1.40	1.37	1.09	1.31	0.81	1.42
BR1	2.59	2.83	1.85	2.43	1.08	2.15
BR2	2.43	2.36	1.51	1.75	1.09	2.04
CTEF	2.87	2.81	2.07	2.66	1.19	1.70

Bold figures denote statistically significant coefficients at the 10% level.

Table 2 Risk and return of mean-variance efficient portfolios, 1990–2001

Universe	Total active	t-value	Asset selection	t-value	Risk index	t-value
RMC	1.98	1.37	0.99	0.86	0.97	1.45
R1000	2.47	2.52	1.85	2.12	0.82	2.13
R2500	7.76	4.37	6.48	3.96	1.61	2.85
R2000	9.68	5.83	8.81	5.57	0.90	2.36

Where RMC = Frank Russell Mid-Cap Universe; R1000 = Frank Russell Largest 1000 Stock Universe; R2000 = Frank Russell Small-Cap Universe; R2500 = Frank Russell Small- and Mid-Cap Universe

Active returns rise as the average stock size declines – a result consistent with the inefficient markets literature summarized in Dimson (1988) and Ziemba (1993). The highest total active returns are found in the Russell 2000 stocks, which are the smallest stocks in the 3000 universe each year (see Table 2). The CTEF tilt variable does produce statistically significant sector exposures in the Russell 2000 stocks (again, see Table 2), just as we previously noted in the Russell 3000 universe.

The CTEF variable produces statistically significant asset selection in the Russell universes, with the exception of the mid-cap universe during the 1990–2001 period. In 2002, Guerard and Mark expanded CTEF testing to worldwide stock markets. We call this Global CTEF variable the Global Analyst Forecasting Efficiency Model – GAFEM. Table 3 reports the GAFEM information coefficients for the 1990–2001 period. It was computed in the same way as the US and non-US CTEF variables. GAFEM was highly statistically significant in all countries at the 10% level. All countries or index constituents were statistically significant at the 5% level, with the exceptions of the Russell 1000 stocks and German stocks during the 1990–2001 time period.

Miller et al. (2014) updated Guerard and Mark (2004) using US-E3 for CTEF and the nine-factor USER Model for the 1979–2009 time period. Miller et al. (2014) reported a “beautiful trade-off curve” and highly statistically significant positive management returns (active returns) and asset selection (stock selection). A slightly edited form is in Guerard et al. (2019). As one increases the portfolio lambda, thereby decreasing the risk aversion level, the number of assets in the portfolio

Table 3 Global Analysts' earnings forecast efficiency, 1990–2001

Universe	Factor	U Return	Q1 Return	Q5 Return	Q1-Q5	Q1 Xs	Batting Avg.	IC	T-Stat
Japan All	CTEF	−2.60	6.99	−9.57	16.56	9.59	77.25	0.067	3.33
UK All	CTEF	11.03	21.35	−0.23	21.58	10.32	73.05	0.073	2.88
Germany All	CTEF	6.27	13.41	−4.85	18.26	7.14	57.49	0.063	1.84
France All	CTEF	14.04	23.71	0.73	22.98	9.67	67.07	0.078	2.19
Australia All	CTEF	12.50	25.43	−1.49	26.92	12.93	72.62	0.085	2.33
US R1000	CTEF	13.95	16.62	9.75	6.87	2.67	59.65	0.032	1.79
US R2000	CTEF	12.58	21.66	−10.53	32.19	9.08	71.93	0.058	5.64

falls, the managed returns rise, and the information ratio rises versus either the Russell 3000 Growth Index or the S&P 500 Index (see Table 4). Asset selection is statistically significant and substantial, exceeding 350 basis points for the 30-year backtest. It generated statistically significant portfolio returns, passing the Markowitz-Xu (1994) Data Mining Corrections test.

Barra re-estimated its risk models, creating GEM2 in 2010, the first global equity model revision in over 20 years, and creating US-E4 and GEM4 in mid-2015 – all under the direction of Jose Menchero.⁶ In Table 5, we summarize the risk and return

⁶Jose Menchero updated the Barra Global Equity Model (GEM) in 2010, see Menchero et al. (2010). Menchero, Morozov, and Shepard estimated the GEM2 Model, which had eight style factors: volatility, momentum, size, value, growth, nonlinear size, liquidity, and leverage. The original Global Equity Model (GEM) Model that was in place when John Guerard became Director of Quantitative Research at McKinley Capital Management, in 2005, had four style factors: size (the logarithm of market capitalization), success (composed of relative strength), value (composed of earnings to price, book to price, dividend yield, and analyst predicted earnings to price), and variability in markets (historical sigma). Menchero et al. (2010) created three sets of factor portfolios. First, simple factor returns, with factors standardized to have cap-weighted mean 0 and standard deviation 1, producing factor-replicating portfolios that go long stocks with positive factor exposures and short stocks with negative factor exposures. Simple factor portfolios can have significant exposures to other factors. Second, pure factor exposures in which every stock has unit exposure to the World factor and exposures of 0 and 1 to country and industries. Pure factor-replicating portfolios go long stocks with positive factor exposures and short stocks with negative factor exposures and have zero exposures to other style factors. The third set of factor portfolios are optimized factor portfolios which have unit exposure to the particular factor and zero exposures to all other factors. The GEM2 momentum factor enhanced wealth in simple and pure factor models portfolios, which dominated the optimized momentum factor portfolio for the 1996–2010 time period. Optimized factor portfolios for size, nonlinear size, liquidity, and leverage underperformed the simple and pure factor portfolios, implying that the impact of these variables weakened owing to country, industry, and style factor interactions. See Menchero et al. (2010) in Guerard (2010) and Lee et al. (2013), Chap. 25, written by Menchero, Morozov, and Guerard.

Table 4 Barra analysis of the USER variable December 1979 – December 2009 Universe: US stocks on Barra Benchmark: S&P 500

	Total Mged	Portf Std	Total Active	Portfolio Tracking	Inform Ratio	t-stat on	Asset	Asset Select	Asset Select	Nb of	Total Transact
λ	<u>Return</u>	<u>Dev</u>	<u>Ret</u>	<u>Error</u>	<u>(IR)</u>	<u>IR</u>	<u>Select</u>	<u>IR</u>	<u>t-stat</u>	<u>Stocks</u>	<u>Costs</u>
0.01	13.69	17.65	7.79	10.05	0.76	2.42	6.75	1.03	3.27	67	2.78
0.02	12.24	17.3	6.34	9.13	0.69	2.19	5.24	0.88	2.77	80	2.74
0.025	12.11	17.34	6.21	9.01	0.56	2.18	5.26	0.89	2.8	84	2.73
0.05	10.05	16.77	4.15	7.51	0.56	1.78	4.02	0.79	2.51	104	2.67
0.09	9.51	16.33	3.61	6.46	0.56	1.78	4.01	0.87	2.75	124	2.65
0.15	8.82	16.01	2.92	5.46	0.52	1.65	3.7	0.88	2.77	147	2.62
0.2	8.14	15.77	2.24	5.16	0.11	1.39	3.26	0.83	2.63	163	2.60

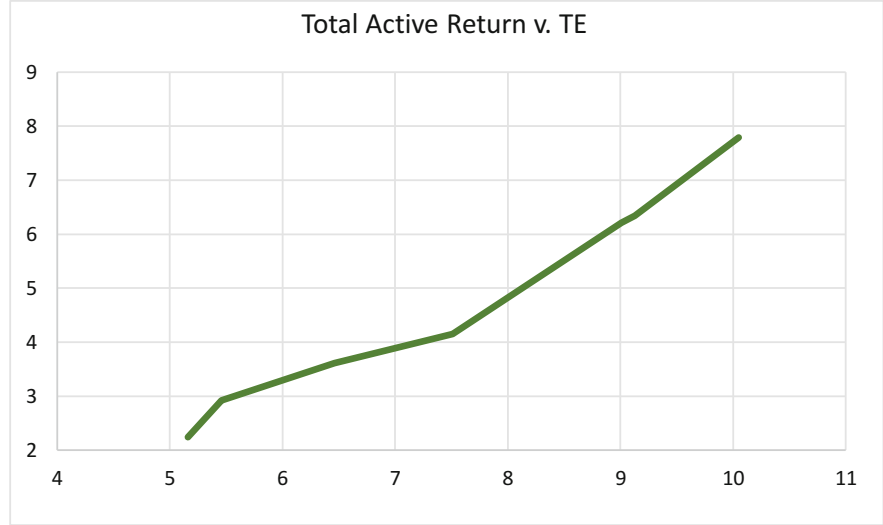
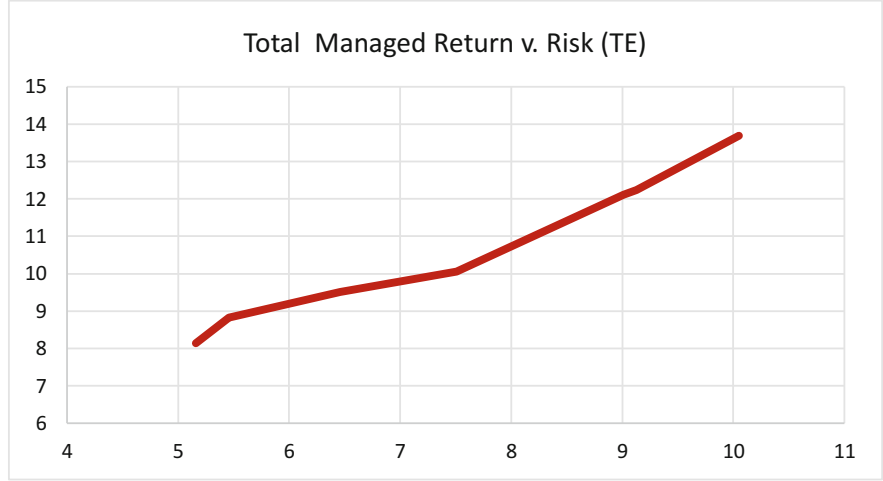


Table 5 Barra US-E4, GEM4. Pure factor portfolios under three distinct regression-weighting schemes. Factor exposures were equal to 1 in all cases. The sample period is from January 1997 through August 2014

Variable	Market Segment	Factor Name	Cap Weight	Root-cap Weight	Equal Weight
Return	USA	Earnings Yield	2.31	2.08	1.81
Risk	USA	Earnings Yield	5.23	3.72	3.22
IR	USA	Earnings Yield	0.44	0.56	0.56
Return	International	Earnings Yield	2.38	2.37	2.35
Risk	International	Earnings Yield	2.53	1.71	1.48
IR	International	Earnings Yield	0.94	1.38	1.59
Return	USA	Momentum	1.37	1.66	1.85
Risk	USA	Momentum	7.08	6.48	6.39
IR	USA	Momentum	0.19	0.26	0.29
Return	International	Momentum	3.30	4.52	5.16
Risk	International	Momentum	5.05	4.18	3.87
IR	International	Momentum	0.65	1.08	1.33

Source: Menchero and Nagy (2017)

characteristics of the various pure factor portfolios. Pure style factor portfolios have unit exposure to a particular style and zero exposure to all other styles and industries. Several points are worth highlighting: For earnings yield, we see that the cap-weighted strategy outperformed the equal-weighted strategy. For momentum, however, the opposite was true. Hence, smaller stocks had more momentum returns relative to risk. We also see that the equal-weighted regressions produced the lowest-volatility portfolios. This was due to excessive concentration in the portfolios constructed using cap-weighted regression. Finally, we note that the International strategies had much higher information ratios than their USA counterparts in both earnings yields and momentum. This was due to the combined effect of higher returns and lower risk.

97.5 The APT Risk Model

John Blin and Steve Bender created APT in 1987. John Blin et al. (1997) and Guerard (2012) demonstrated the effectiveness of the APT (later SunGard APT, and now FIS APT) systems in portfolio construction and management. Below we start with an overview of the APT approach to portfolio construction.

The estimation of security weights, w , in a portfolio is the primary calculation of Markowitz's portfolio management approach. In APT, the security weight is the proportion of the portfolio's market value invested in the individual security:

$$w_s = \frac{MV_s}{MV_p} \quad (15)$$

where

w_s = portfolio weight in security

MV_s = value of security s within the portfolio

MV_p = the total market value of the portfolio

The active weight of the security is calculated by subtracting the security weight in the (index) benchmark b , from the security weight in the portfolio p

$$w_{sp} - w_{sb} \quad (16)$$

The APT Analytics Guide (2005) builds upon the foundations of the Blin and Bender APT system. See Blin et al. (1997). The following analysis draws upon APT analytics. Volatility can be broken down into systematic and specific risk:

$$\sigma_p^2 = \sigma_{\beta p}^2 + \sigma_{ep}^2 \quad (17)$$

where

σ_p = total portfolio volatility

$\sigma_{\beta p}^2$ = systematic portfolio volatility

σ_{ep} = specific portfolio volatility

Blin and Bender created a multi-factor risk model within their APT risk model based on forecast volatility:

$$\sigma_p = \sqrt{52 * \left(\sum_{c=1}^C \sum_{i=1}^S w_i \beta_{i,c} \right)^2 + 52 * \sum_{i=1}^S w_i^2 \varepsilon_i^2} \quad (18)$$

where

σ_p = forecast volatility of annual portfolio return

C = number of statistical components in the risk model

w_i = portfolio weight in security i

$\beta_{i,c}$ = loading (beta) of security i on risk component c

$\varepsilon_{i, week}$ = weekly specific volatility of security i

The multiplier 52 is the annualization factor because there are 52 weeks per year and the calculations use weekly returns.

Tracking error is a measure of volatility applied to the active return of funds (or portfolio strategies) against a benchmark, which is often an index. Portfolio tracking error is defined as the standard deviation of the portfolio return *less* the benchmark return over 1 year.

$$\sigma_{te} = \sqrt{E((r_p - r_b) - E(r_p - r_b))^2} \quad (19)$$

where

σ_{te} = annualized tracking error

r_p = actual portfolio annual return

r_b = actual benchmark annual return

The systematic tracking error of a portfolio is a forecast of the portfolio's active annual return as a function of the securities' returns associated with APT risk model components. Portfolio-specific tracking error can be written as a forecast of the annual portfolio active return associated with each security's specific behavior:

$$\sigma_{ste} = \sqrt{52 \sum_{i=1}^{N_s} (w_{i,p} - w_{i,b})^2 \varepsilon_{i,week}^2 R_p} \quad (20)$$

The APT calculated portfolio tracking error versus a benchmark is:

$$\sigma_{p-b} = \sqrt{52 (w_p - w_b)^T (b^T b + \varepsilon^T \varepsilon) (w_p - w_b)}, \quad (21)$$

where

σ_{p-b} = forecast tracking error

$b = a$ ($N_c \times N_s$) matrix of component loadings; N_c components in the model and N_s securities in the portfolio

ε = a diagonal matrix $N_c \times N_s$ of the specific loadings

$w_p - w_b$ = the (N_s -dimensional) security active-weights vector

Thus the forecast tracking error is:

$$\sigma_{p-b} = \sqrt{52 \left[\left(\sum_{c=1}^{N_c} \sum_{s=1}^{N_s} (w_{sp} - w_{sb}) b_{sc} \right)^2 + \sum_{s=1}^{N_s} (w_{sp} - w_{sb})^2 \varepsilon_s^2 \right]} \quad (22)$$

where

$w_{sp} - w_{sb}$ = portfolio active weights in security s

b_{sc} = the loading of security s on component c

ε_s = weekly specific volatility of security s

The marginal tracking error is a measure of the sensitivity of the tracking error of an active portfolio to changes in the active weight of a specific security.

$$\partial_s [\sigma_{p-b}] = \frac{\partial \sigma_{p-b}}{\partial w_{s,p-b}} \quad (23)$$

where $w_{s,p-b}$ = active portfolio weights in security s .

The APT calculated contribution-to-risk of a security is:

$$\Delta[\sigma_p] = \frac{52(b^T b + s^T s)w}{\sqrt{52(b^T b + s^T s)w}} \quad (24)$$

where $\Delta[\sigma_p] = A$ ($N_s - \text{dimensional}$) vector of contribution to volatility for securities in the portfolio with N_s securities.

The portfolio Value-at-Risk (VaR) is a measure of the distribution of expected outcomes. If one is concerned with a 95% confidence level α and a 30-day time horizon T , then the 95%, 30-day VaR of the portfolio is the minimum we would expect to lose and is characterized by

$$P[V_T < V_0 - v(\alpha, T)] = 1 - \alpha \quad (25)$$

where

$v(\alpha, T)$ = VaR at the confidence interval α for time T

V_T = portfolio value at time T

α = required confidence level

If the portfolio returns are normally distributed:

$$v_s(\alpha, T) = V_0 - E(V_T) - V_0 \phi^{-1}(\alpha) \sigma_p^T \quad (26)$$

where

$v_s(\alpha, T)$ = Gaussian VaR

σ_p^T = forecast portfolio volatility

$\phi^{-1}(\alpha)$ = inverse cumulative normal distributed function

then the total VaR is:

$$v_c(\alpha, T) = V_0 \sqrt{\frac{\alpha}{1-\alpha} \left(\sigma_{\beta p}^T + \phi^{-1}(\alpha) \sigma_{sp}^T \right)^2} \quad (27)$$

The Tracking-at-Risk (TaR) is a measure of portfolio risk estimating the magnitude that the portfolio return may deviate from the benchmark return over time. The maximum deviation of a portfolio return over the time horizon T at given confidence level α is

$$v_{p-b}(\alpha, T) = \sqrt{\left(\frac{1}{\sqrt{1-\alpha}} \sigma_{\beta, p-b}^T \right)^2 + \phi^{-1}\left(\frac{1+\alpha}{2}\right) \sigma_{s, p-b}^2} \quad (28)$$

How relevant is APT and Markowitz Mean-Variance analysis? We shall see shortly. Spoiler alert: Harry Markowitz (1952 and 1959) produces statistically significant portfolio Active Returns and Specific Returns (stock selection) when used with a statistically significant tilt variable. An excellent and fairly recent book

on portfolio risk is Connor et al. (2010). Korajczyk provides a comprehensive bibliography on APT and multi-factor risk models at the following link: <https://www.kellogg.northwestern.edu/faculty/korajczyk/htm/aptlist.htm>.

The reader is also referred to Lee et al. (2013), Chap. 13, for a very useful discussion of multi-index and APT models.

Blin et al. (1997) used an estimated 20-factor beta model of covariances based on 3.5 years of weekly stock return data. The Blin and Bender Arbitrage Pricing Theory (APT) Model followed the Ross factor modeling theory, but Blin and Bender estimated betas from at least 20 orthogonal factors. Empirical support is reported in Guerard, Chettiappan, and Xu (2010), Guerard (2012), and Guerard, Markowitz, and Xu (2015) for the application of mean-variance and mean-variance Tracking-at-Risk optimization techniques.

It is well known that as one raises a portfolio's lambda, the expected return of the portfolio rises and the number of securities in that optimal portfolio falls. See Blin et al. (1997). Lambda, a measure of risk aversion (the inverse of the risk-aversion acceptance level of the Barra system), is a decision variable to control the optimal number of securities in a portfolio. As lambda increases, the number of stocks in the portfolio falls and the Sharpe ratio rises. A more aggressive, growth-maximizing asset manager will prefer a higher lambda.

Although the Markowitz Mean-Variance (M59) portfolio construction and management analysis is six decades old, it performs well in terms of maximizing the Sharpe ratio, geometric mean, and information ratio relative to newer approaches. See Table 6 for a streamlined version of the Markowitz Mean-Variance and Mean-Variance Tracking-at-Risk (MVTaR) simulations of Guerard, Markowitz, and Xu (2015). A lambda of 200, an aggressive risk-tolerance parameter, is sufficient in the USA and global markets to maximize the geometric mean, Sharpe ratio, and information ratio for the MCM variables.⁷ A lambda of 200 MVTaR GLER simulation outperforms a portfolio based on Markowitz's 1959 *Portfolio Selection* (M59) on an information ratio (IR) basis: 0.97 versus 0.85. MVTaR does not dominate the traditional Markowitz Mean-Variance model, but it does produce lower tracking errors and higher IRs. Now, to illustrate an enhanced-index-tracking (EIT) portfolio strategy, we use a much lower lambda of 5.⁸ Guerard, Markowitz, and Xu (2015) updated Guerard and Mark (2004),

⁷Guerard and Schwartz (2007), Chap. 15, identified a lambda of 200 as the Sharpe-Ratio-maximizing lambda. Further support can be found in Guerard, Rachev, and Shao (2013), reprinted in Guerard and Ziemba (2020).

⁸A very important result of APT portfolio modeling is reported in Guerard, Krauklis, and Kumar (2012) for the application of mean-variance, enhanced index tracking and Tracking-at-Risk optimization techniques for the 1997–2009 time period. We can optimize the US stocks to create portfolios in the Russell 3000 universe for the 12-year backtest period, 1997–2009, where the absolute value of the stock weight cannot deviate by more than 2%. Guerard, Krauklis, and Kumar labeled these as (1) an EIT portfolio or (2) an EAW2 (equal active weight of 2%) portfolio. For a lambda of 200, 99 stocks are needed in the mean-variance portfolios and 108 stocks in the EAW2 portfolios. For a lambda of 10, we need 159 stocks in the mean-variance portfolios and 158 stocks in the EAW2 portfolios. An index-hugging portfolio – one designed by a manager to avoid taking active bets – would use a lambda of 10. An even less aggressive manager would need 161 stocks in an EAW1 portfolio for a lambda of 10, whereas a manager using a Sharpe-Ratio-maximizing lambda of 100 would need 131 stocks.

Table 6 Efficient frontier of the global stock selection model, GLER, with various portfolio optimization techniques and universes: (1) the global I/B/E/S stocks with at least two analysts, 1999–2011 and (2) the largest 7,500 global stocks (in USD)

APT Risk Model and Optimization System

Earnings	Mean-					
Model or	Variance		Annualized	Sharpe	Information	Tracking
Component	Methodology	Lambda	Return	Ratio	Ratio	Error
GLER	M59	1000	15.84	0.590	0.78	13.11
		500	16.34	0.590	0.82	12.08
		200	16.37	0.610	0.85	12.68
		100	15.90	0.580	0.81	12.66
		5	10.11	0.440	0.51	8.81
Benchmark			5.59	0.240		
GLER	MVTaR	1000	16.10	0.660	0.94	11.18
		500	15.91	0.651	0.90	11.44
		200	16.09	0.691	0.97	10.83
		100	14.18	0.591	0.77	11.23
		5	8.51	0.344	0.33	8.75

Constraints:

- 1. Turnover is 8% (buys) monthly;
- 2. Long-only holdings (a 4% maximum stock weight); and
- 3. Threshold positions of 35 basis points.

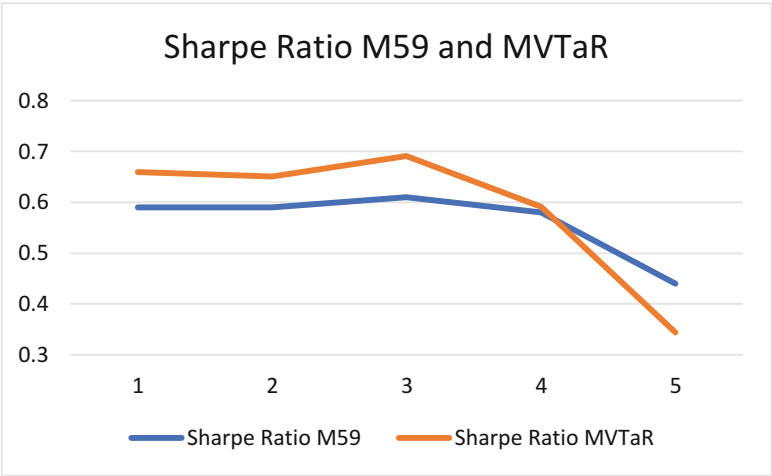
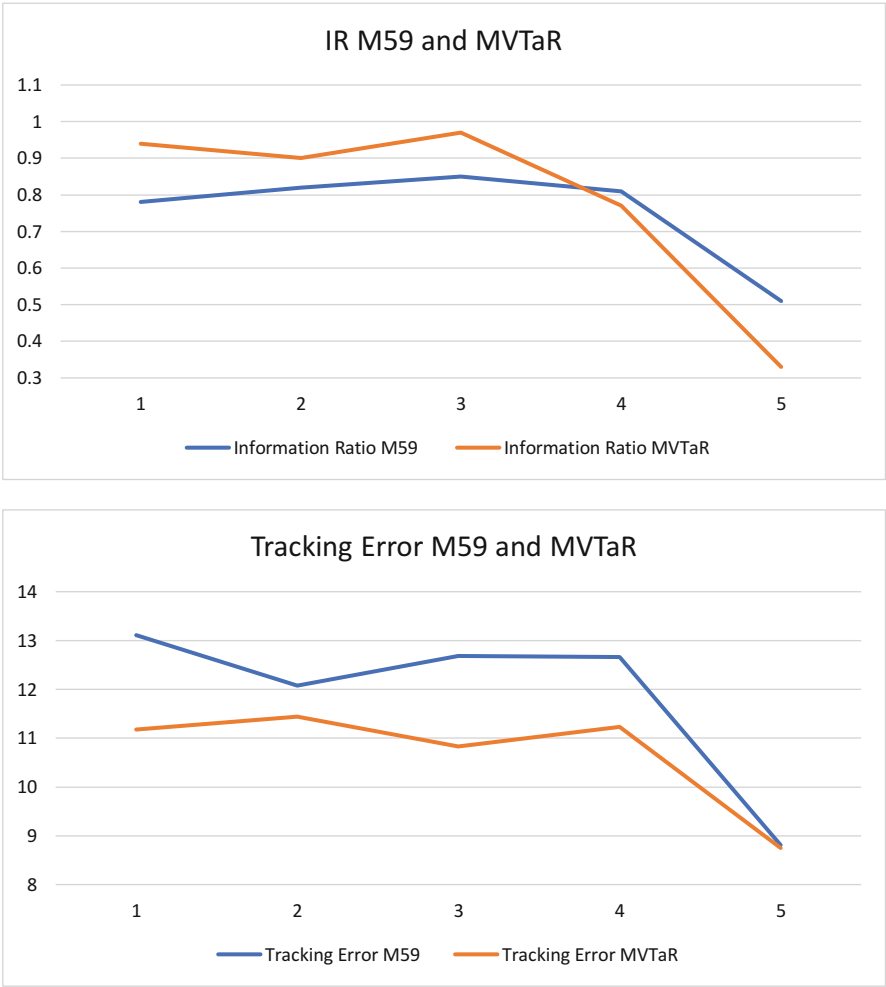


Table 6 (continued)



highlighting further evidence of the continued statistical significance of earnings forecasting analysis. In the world of applied investment research, the authors of this chapter believe that all three levels of portfolio testing must be passed to achieve statistical significance.⁹

⁹Guerard brought the Blin and Bender APT system to MCM in August 2005. The APT system produced real-time portfolios that generated statistically significant Active Returns and Specific Returns. See Guerard, Deng, Gillam, Markowitz, and Xu (2019). The real-time Specific Returns exceeded 300 basis points, very consistent with the model’s backtests. See Guerard and Markowitz (2018). Full disclosure: Guerard worked with APT as a client from 1988 until early 2008, when Blin and Bender sold the company to SunGard.

97.6 The Axioma Statistical and Fundamental Risk Models

The Axioma Robust Risk Model is a multi-factor model, in the tradition of the Barra model and Eq. (20) above.¹⁰ Axioma offers both the US and World Fundamental and Statistical Risk Models. The Axioma Risk Models use several statistical techniques to efficiently estimate factors. The ordinary least squares residuals (OLS) are not homoscedastic; that is, when one minimizes the sum of the squared residuals to estimate factors using OLS, one finds that large assets exhibit lower volatility than smaller assets. A constant variance of returns is not found. Axioma uses a weighted least squares (WLS) regression, which scales the asset residual by the square root of the asset market capitalization (to serve as a proxy for the inverse of the residual variance). Robust regression, using the Huber M-Estimator, addresses the issue and problem of outliers. Asymptotic principal components analysis (PCA) is used to estimate the statistical risk factors. A subset of assets is used to estimate the factor returns, and the exposures and factor returns are applied to the other assets.

The Axioma World-Wide Equity Risk Factor Model (AX-WW2.1) seeks to forecast medium-term horizon risk, or risk 3–6 months ahead. The Equity Risk Factor, or Fundamental, Model uses nine style factors: exchange rate sensitivity, growth (historical earnings and sales growth), leverage (debt-to-assets), liquidity (1-month trading volume divided by market capitalization), medium-term momentum (cumulative returns of the past year, excluding the previous month), short-term momentum (last month return), size (natural logarithm of issuer market capitalization), value (book-to-price and earnings-to-price ratios), and volatility (3-month average of absolute returns divided by cross-sectional standard deviation). The Axioma Statistical Risk Model estimates 15 factors using principal components.

97.7 The MCM Horse Races: Who Has the Better Risk Models for Growth Variables?

In 2011, McKinley Capital Management, LLC (MCM) initiated a “Horse Race” testing procedure to find out whether all risk models and optimizers were created equal. Bottom line: They are not. The results appeared in the Spring 2012 *Journal of Investing*, a special edition devoted to Quantitative Risk Management. It featured articles by Markowitz; Tsuchida, Zhou, and Rachev (the Cognity system of FinAnalytics); Wormald and van Der Merwe (APT); Saxena and Stubbs (Axioma); and Guerard, Xu, and Gültekin (APT, Barra, and Cognity). In 2011, APT and

¹⁰ *Axioma Risk Model Handbook*, January 2010.

Table 7 APT and Axioma Statistical Risk Models, US stocks, 1998–2009

Assumptions and constraints							
Universe: All US stocks in the WRDS database							
Time period: January 1998 to December 2009							
Position constraint							
Minimum position = 0.0							
Maximum position = 4.0							
Position threshold holding = 0.35%							
Turnover = 8% monthly							
Transaction cost: 1.25% each way							
Risk model:							
Axioma – Axioma Statistical Model							
APT – APT Statistical Model							
Risk setting:							
Axioma – Tracking error target of 8%							
APT – Lambda of 200							
Model:	Expected return model: Regression-weighted EP BP CP SP REP RBP RCP RSP CTEF PM						
Benchmark: R3000G							
	SHR	IR	TE	SD	AR	Asset Selection	T-Stat
Axioma	0.177	0.69	7.78	21.92	7.06	6.00	5.09
APT	0.117	0.40	10.29	23.02	5.87	6.77	3.92
R3000G	–0.074			19.73	1.73		

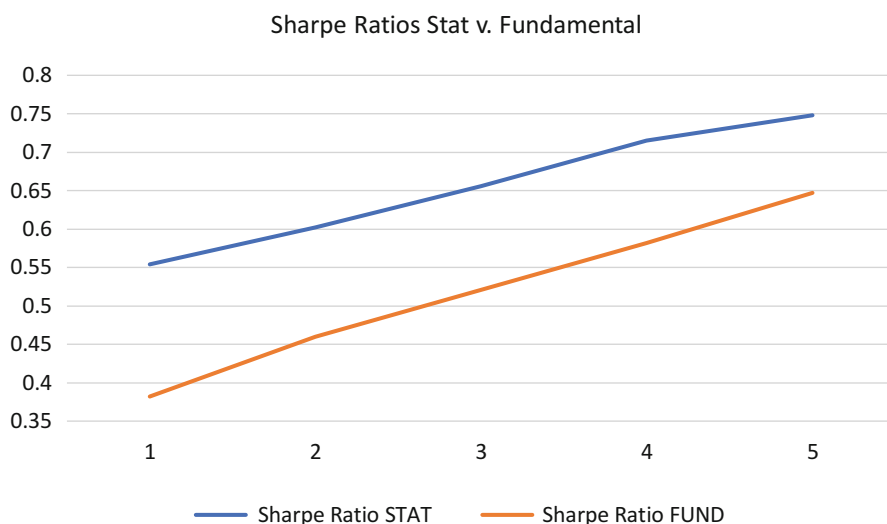
Where SHR = Sharpe ratio, IR = Information ratio, TE = Tracking error, SD = Portfolio standard deviation, AR = Average annual return, Asset selection = Barra-attributed stock selection

Axioma were the winners among five risk models and optimization systems using the MCM US “Public Models” CTEF and USER. As Table 7 shows, both APT and Axioma optimization systems produced highly statistically significant asset selection, exceeding 600 basis points annually (the Barra attribution system was used as a judge).

What factor contributions of the Axioma portfolio were rewarded by the market for 1998–2009? The answer is small size (the exposure of –1.17 produced 329 basis points of return with a corresponding t-statistic of 3.16) and earnings yield (the exposure of 0.09 produced 41 basis points of return with a corresponding t-statistic of 1.85, statistically significant at the 10% level). Growth, leverage, and yield produced negative statistically significant factor returns in the Barra USER attributions for 1998–2009. APT and Axioma both had statistically based risk models; Axioma and Barra had fundamentally based risk models.

In 2012, MCM ran a second Horse Race competition using a two-I/B/E/S analyst, 7500-largest market-capitalized global universe. The Axioma Statistical Risk Model produced higher geometric means, Sharpe ratios, and information ratios using GLER data than its Fundamental (WW21AxiomaMH) Model generated.

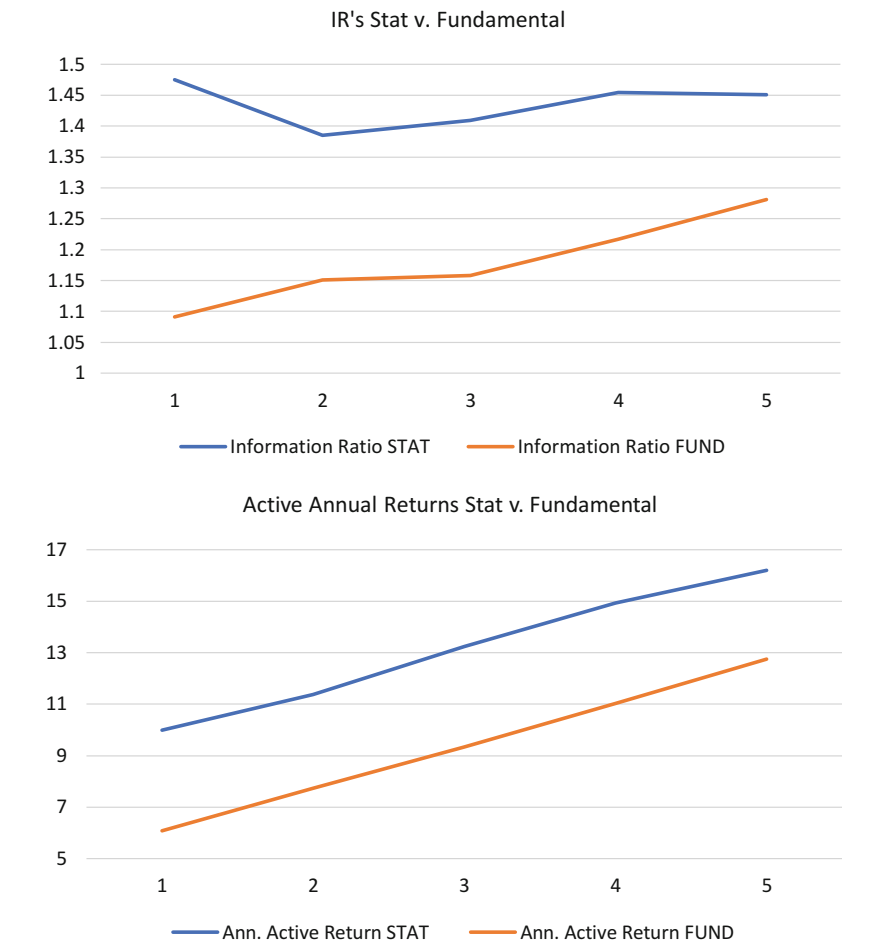
Guerard, Markowitz, and Xu (2014 and 2015) reported efficient frontiers using both of the Axioma Risk Models. As noted above, the Axioma Statistical Risk Model, which the authors denoted as “STAT,” produced geometric means, Sharpe ratios, and information ratios that were higher than those generated by the Axioma Fundamental Risk Model, denoted as “FUND.” The geometric means and Sharpe ratios increase with the targeted tracking errors; however, the information ratios are higher in the lower tracking error range of 3–6%, with at least 200 stocks, on average, in the optimal portfolios. See Table 8. The Guerard et al. (2013, 2015) studies assumed 150 basis points, each way, of transaction costs. The use of ITG cost curves produced about 115–125 basis points of transaction costs, well under the assumed costs. The Guerard et al. studies also used the SunGard APT statistical model, which produced statistically significant asset selection in the US and global portfolios. In the MCM Horse Races, 2010–2017, the most consistent finding is the superior performance of statistical models.¹¹

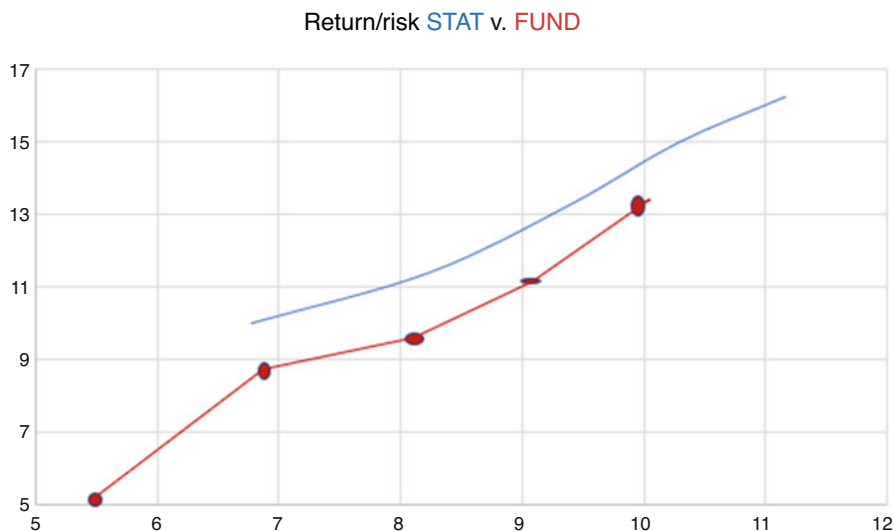


¹¹ In 2017, MCM ran its third Horse Race using Axioma with integrated ITG transactions costs, in the MSCI Non-US (XUS), Global (GL), and Emerging Markets (EM) universes. The XUS, GL, and EM reported results show much higher Active and Specific Returns in non-US markets than in the US market. International markets are more inefficient than the US markets, as was reported in Guerard and Mark (2020).

Table 8 Axioma analysis of the FactSet global 7500-stock data Universe, AWCG Universe, January 1999 – December 2011. Transactions costs: 150 basis points each way, respectively

Return model	Risk model	Tracking error	Sharpe ratio	Information ratio	Ann. active return	Ann. active risk	N
GLER	STAT	4	0.554	1.475	9.99	6.78	144
		5	0.602	1.385	11.38	8.24	110
		6	0.656	1.409	13.25	9.40	87
		7	0.715	1.454	14.94	10.28	70
		8	0.748	1.451	16.20	11.16	58
	FUND	4	0.382	1.091	6.08	5.57	163
		5	0.460	1.151	7.73	6.72	129
		6	0.521	1.158	9.33	8.06	104
		7	0.582	1.217	11.02	9.06	83
		8	0.647	1.281	12.75	9.95	71





97.8 Summary and Conclusions

Multi-factor risk models may be used to analyze a stock selection or evaluate a variable. Blin et al. (1997), Guerard et al. (2014, 2015), Guerard and Mark (2004, 2020), and Blin et al. (1997) reported that properly constructed risk models created portfolios in which tilt variables that have statistically significant ICs produced statistically significant Active Returns. One must remove the market effect and extra-market covariances to properly estimate how much a variable contributes to the creation of efficient portfolios.

It is useful to see how quantitative analysis was developed in investment research and analysis. Harry Markowitz, Bill Sharpe, Jan Mossin, and Jack Treynor pioneered capital market equilibrium and the creation and estimation of the Capital Asset Pricing Model. In the 1970s, multi-factor risk models were developed and estimated by Barr Rosenberg, Andrew Rudd, and their colleagues at Barra; John Blin and Steve Bender at APT; and Sebastian Ceria at Axioma. Recent research and ongoing implementation by APT and Axioma have further refined portfolio optimization. MCM developed its Horse Race risk model philosophy to assess the most effective risk models for maximizing the geometric means, Sharpe ratios, and information ratios. APT and Axioma have reigned supreme in those tests. The integration of Axioma and AAF and ITG transactions costs on the FactSet Investment have led MCM to implement an Axioma portfolio.

Acknowledgments The authors would like to thank research conversations with Doug Martin, Bijan Beheshti, and Sebastian Ceria. Any errors remaining are the sole responsibility of the authors.

Appendix 1: US-E3 Descriptor Definitions

This appendix gives detailed definitions of the descriptors that underlie the risk indices in US-E3. The method of combining these descriptors into risk indices is proprietary to Barra.

1. Volatility

(i) *BTSg: Beta times sigma*

This is computed as $\sqrt{\beta\sigma_e}$, where β is the historical beta and σ_e is the historical residual standard deviation. If β is negative, then the descriptor is set equal to zero.

(ii) *DASTD: Daily standard deviation*

This is computed as:

$$\sqrt{N_{\text{days}} \left[\sum_{t=1}^T w_t r_t^2 \right]}$$

where r_t is the return over day t , w_t is the weight for day t , T is the number of days of historical returns data used to compute this descriptor (we set this to 65 days), and N_{days} is the number of trading days in a month (we set this to 23).

(iii) *HILO: Ratio of high price to low price over the last month*

This is calculated as:

$$\log \left(\frac{P_H}{P_L} \right)$$

where P_H and P_L are the maximum price and minimum price attained over the last 1 month.

(iv) *LPRI: Log of stock price*

This is the log of the stock price at the end of last month.

(v) *CMRA: Cumulative range*

Let Z_t be defined as follows:

$$Z_t = \sum_{s=1}^t \log(1 + r_{i,s}) - \sum_{s=1}^t \log(1 + r_{f,s})$$

where $r_{i,s}$ is the return on stock i in month s , and $r_{f,s}$ is the risk-free rate for month s . In other words, Z_t is the cumulative return of the stock over the risk-free rate at the end of month t . Define $Z_{\max}$ and $Z_{\min}$ as the maximum and minimum values of Z_t over the last 12 months. CMRA is computed as:

$$\log \left(\frac{1 + Z_{\max}}{a + Z_{\min}} \right)$$

(vi) *VOLBT: Sensitivity of changes in trading volume to changes in aggregate trading volume*

This may be estimated by the following regression:

$$\frac{\Delta V_{i,t}}{N_{i,t}} = a + b \frac{\Delta V_{M,t}}{N_{m,t}} + \xi_{i,t}$$

where $\Delta V_{i,t}$ is the change in share volume of stock I from week t-1 to week t, $N_{i,t}$ is the average number of shares outstanding for stock I at the beginning of week t-1 and week t, $\Delta V_{M,t}$ is the change in volume on the aggregate market from week t-1 to week t, and $N_{M,t}$ is the average number of shares outstanding for the aggregate market at the beginning of week t-1 and week t.

(vii) *SERDP: Serial dependence*

This measure is designed to capture serial dependence in residuals from the market model regressions. It is computed as follows:

$$\text{SERDP} = \frac{\frac{1}{T-2} \sum_{t=3}^T (e_t + e_{t-1} + e_{t-2})^2}{\frac{1}{T-2} \sum_{t=3}^T (e_t^2 + e_{t-1}^2 + e_{t-2}^2)}$$

where e_t is the residual from the market model regression in month t , and T is the number of months over which this regression is run (typically, $T = 60$ months).

(viii) *OPSTD: Option-implied standard deviation*

This descriptor is computed as the implied standard deviation from the Black-Scholes option pricing formula using the price on the closest to at-the-money call option that trades on the underlying stock.

2. Momentum

(i) *RSTR: Relative strength*

This is computed as the cumulative excess return (using continuously compounded monthly returns) over the last 12 months – i.e.,

$$\text{RSTR} \rightleftharpoons = \sum_{t=1}^T \log(1 + r_{i,t}) - \sum_{t=1}^T \log(1 + r_{f,t})$$

where $r_{i,t}$ is the arithmetic return of the stock in month I , and $r_{f,t}$ is the arithmetic risk-free rate for month i . This measure is usually computed over the last 1 year – i.e., T is set equal to 12 months.

(ii) *HALPHA: Historical alpha*

This descriptor is equal to the alpha term (i.e., the intercept term) from a 60-month regression of the stock's excess returns on the S&P 500 excess returns.

3. Size

(i) *LNCAP: Log of market capitalization*

This descriptor is computed as the log of the market capitalization of equity (price times number of shares outstanding) for the company.

4. Size nonlinearity

(i) *LCAPCB: Cube of the log of market capitalization*

This risk index is computed as the cube of the normalized log of market capitalization.

5. Trading activity

i) *STOA: Share turnover over the last year*

STOA is the annualized share turnover rate using data from the last 12 months – i.e., it is equal to $V_{\text{ann}} / \bar{N}_{\text{out}}$, where V_{ann} is the total trading volume (in number of shares) over the last 12 months and $\bar{N}_{\text{out}}$ is the average number of shares outstanding over the previous 12 months (i.e., it is equal to the average value of the number of shares outstanding at the beginning of each month over the previous 12 months).

ii) *STOQ: Share turnover over the last quarter*

This is computed as the annualized share turnover rate using data from the most recent quarter. Let V_q be the total trading volume (in number of shares) over the most recent quarter and let $\bar{N}_{\text{out}}$ be the average number of shares outstanding over the period (i.e., $\bar{N}_{\text{out}}$ is equal to the average value of the number of shares outstanding at the beginning of each month over the previous 3 months). Then, STOQ is computed as $4V_q / \bar{N}_{\text{out}}$.

iii) *STOM: Share turnover over the last month*

This is computed as the share turnover rate using data from the most recent month – i.e., it is equal to the number of shares traded last month divided by the number of shares outstanding at the beginning of the month.

iv) *STO5: Share turnover over the last 5 years*

This is equal to the annualized share turnover rate using data from the last 60 months. In symbols, STO5 is given by:

$$\text{STO5} = \frac{12 \left[\frac{1}{T} \sum_{s=1}^T V_s \right]}{\sigma_\varepsilon}$$

where V_s is equal to the total trading volume in months, and $\bar{N}_{\text{out}}$ is the average number of shares outstanding over the last 60 months.

v) *FSPLIT: Indicator for forward split*

This descriptor is a 0–1 indicator variable to capture the occurrence of forward splits in the company's stock over the last 2 years.

vi) *VLVR: Volume to variance*

This measure is calculated as follows:

$$\text{VLVR} = \log \frac{\frac{12}{T} \left[\sum_{s=1}^T V_s P_s \right]}{\sigma_\varepsilon}$$

where V_s equals the number of shares traded in month s , P_s is the closing price of the stock at the end of month s , and σ_ε is the estimated residual standard deviation. The sum in the numerator is computed over the last 12 months.

6. Growth

(i) *PAYO: Payout ratio over 5 years*

This measure is computed as follows:

$$\text{PAYO} = \frac{\frac{1}{T} \sum_{t=1}^T D_t}{\frac{1}{T} \sum_{t=1}^T E_t}$$

where D_t is the aggregate dividend paid out in year t and E_t is the total earnings available for common shareholders in year t . This descriptor is computed using the last 5 years of data on dividends and earnings.

(ii) *VCAP: Variability in capital structure*

This descriptor is measured as follows:

$$\text{VCAP} = \frac{\frac{1}{T-1} \sum_{t=2}^T (|N_{t-1} - N_t| P_{t-1} + |LD_{t-1} - LD_t| + |PE_t - PE_{t-1}|)}{CE_T + LD_T + PE_T}$$

where N_{t-1} is the number of shares outstanding at the end of time $t-1$; P_{t-1} is the price per share at the end of time $t-1$; LD_{t-1} is the book value of long-term debt at the end of time period $t-1$; PE_{t-1} is the book value of preferred equity at the end of time period $t-1$; and $CE_T + LD_T + PE_T$ are the book values of common equity, long-term debt, and preferred equity as of the most recent fiscal year.

(iii) *AGRO: Growth rate in total assets*

To compute this descriptor, the following regression is run:

$$TA_{it} = a + bt + \xi_{it}$$

where TA_{it} is the total assets of the company as of the end of year t , and the regression is run for the period $= 1, \dots, 5$. AGRO is computed as follows:

$$AGRO = \frac{b}{\frac{1}{T} \sum_{t=1}^T TA_{it}}$$

where the denominator average is computed over all the data used in the regression.

(iv) *EGRO: Earnings growth rate over last 5 years*

First, the following regression is run:

$$EPS_t = a + bt + \xi_t$$

where EPS_t is the earnings per share for year t . This regression is run for the period $t = 1, \dots, 5$. EGRO is computed as follows:

$$EGRO = \frac{b}{\frac{1}{T} \sum_{t=1}^T EPS_t}$$

(v) *EGIBS: Analyst-predicted earnings growth*

This is computed as follows:

$$EGIBS = \frac{(EARN - EPS)}{(EARN \leftrightarrow + \leftrightarrow EPS)/2}$$

where EARN is a weighted average of the median earnings predictions by analysts for the current year and next year, and EPS is the sum of the four most recent quarterly earnings per share.

(vi) *DELE: Recent earnings change*

This is a measure of recent earnings growth and is measured as follows:

$$\text{DELE} = \frac{(\text{EPS}_t - \text{EPS}_{t-1})}{(\text{EPS}_t + \text{EPS}_{t-1})/2}$$

where EPS_t is the earnings per share for the most recent year, and EPS_{t-1} is the earnings per share for the previous year. We set this to missing if the denominator is nonpositive.

7. Earnings Yield

(i) *EPIBS: Analyst-predicted earnings-to-price*

This is computed as the weighted average of analysts' median predicted earnings for the current fiscal year and next fiscal year divided by the most recent price.

(ii) *ETOP: Trailing annual earnings-to-price*

This is computed as the sum of the four most recent quarterly earnings per share divided by the most recent price.

(iii) *ETP5: Historical earnings-to-price*

This is computed as follows:

$$\text{ETP5} = \frac{\frac{1}{T} \sum_{t=1}^T \text{EPS}_t}{\frac{1}{T} \sum_{t=1}^T P_t}$$

where EPS_t is equal to the earnings per share over year t , and P_t is equal to the closing price per share at the end of year t .

8. Value

(i) *BTOP: Book-to-price ratio*

This is the book value of common equity as of the most recent fiscal year-end divided by the most recent value of the market capitalization of the equity.

9. Earnings Variability

(i) *VERN: Variability in earnings*

This measure is computed as follows:

$$\text{VERN} = \frac{\left(\frac{1}{T-1} \sum_{t=1}^T (E_t - \bar{E})^2 \right)^{1/2}}{\frac{1}{T} \sum_{t=1}^T E_t}$$

where E_t is the earnings at time t ($t = 1, \dots, 5$), and $\bar{E}$ is the average earnings over the last 5 years. VERN is the coefficient of variation of earnings.

(ii) *VFLO: Variability in cash flows*

This measure is computed as the coefficient of variation of cash flow using data over the last 5 years – i.e., it is computed in an identical manner to VERN, with cash flow being used in place of earnings. Cash flow is computed as earnings plus depreciation plus deferred taxes.

(iii) *EXTE: Extraordinary items in earnings*

This is computed as follows:

$$EXTE = \frac{\frac{1}{T} \sum_{t=1}^T |EX_t + NRI_t|}{\frac{1}{T} \sum_{t=1}^T E_t}$$

where EX_t is the value of extraordinary items and discontinued operations, NRI_t is the value of nonoperating income, and E_t is the earnings available to common before extraordinary items. The descriptor uses data over the last 5 years.

(iv) *SPIBS: Standard deviation of analysts' prediction to price*

This is computed as the weighted average of the standard deviation of IBES analysts' forecasts of the firm's earnings per share for the current fiscal year and next fiscal year divided by the most recent price.

10. Leverage

(i) *MLEV: Market leverage*

This measure is computed as follows:

$$MLEV = \frac{ME_t + PE_t + LD_t}{ME_t}$$

where ME_t is the market value of common equity, PE_t is the book value of preferred equity, and LD_t is the book value of long-term debt. The value of preferred equity and long-term debt are as of the end of the most recent fiscal year. The market value of equity is computed using the most recent month's closing price of the stock.

(ii) *BLEV: Book leverage*

This measure is computed as follows:

$$BLEV = \frac{CEQ_t + PE_t + LD_t}{CEQ_t}$$

where CEQ_t is the book value of common equity, PE_t is the book value of preferred equity, and LD_t is the book value of the long-term debt. All values are as of the end of the most recent fiscal year.

(iii) *DTOA: Debt-to-assets ratio*

This ratio is computed as follows:

$$DTOA = \frac{LD_t + DCL_t}{TA_t}$$

where LD_t is the book value of long-term debt, DCL_t is the value of debt in current liabilities, and TA_t is the book value of total assets. All values are as of the end of the most recent fiscal year.

(iv) *SNRRT: Senior debt rating*

This descriptor is constructed as a multi-level indicator variable of the debt rating of a company.

11. Currency Sensitivity

(i) *CURSENS: Exposure to foreign currencies*

To construct this descriptor, the following regression is run:

$$r_{it} = \alpha_1 + \beta_1 r_{mt} + \varepsilon_{it}$$

where r_{it} is the excess return on the stock and r_{mt} is the excess return on the S&P 500 Index. Let ε_{it} denote the residual returns from this regression. These residual returns are in turn regressed against the contemporaneous and lagged returns on a basket of foreign currencies, as follows:

$$\varepsilon_{it} = c_1 + \gamma_{11}(FX)_t + \gamma_{12}(FX)_{t-1} + \gamma_{13}(FX)_{t-2} + \mu_{it}$$

where ε_{it} is the residual return on stock I , $(FX)_t$ is the return on an index of foreign currencies over month t , $(FX)_{t-1}$ is the return on the same index of foreign currencies over month $t-1$, and $(FX)_{t-2}$ is the return on the same index over month $t-2$. The risk index is computed as the sum of the slope coefficients γ_{11} , γ_{12} , and γ_{13} – i.e., $CURSENS = \gamma_{11} + \gamma_{12} + \gamma_{13}$.

12. Dividend Yield

(i) *P_DYLD: Predicted dividend yield*

This descriptor uses the last four quarterly dividends paid out by the company along with the returns on the company's stock and future dividend announcements made by the company to come up with a Barra-predicted dividend yield.

13. Non-estimation Universe Indicator

(i) *NONESTU: Indicator for firms outside US-E3 estimation universe*

This is a 0–1 indicator variable: It is equal to 0 if the company is in the BARRA estimation universe and equal to 1 if the company is outside the BARRA estimation universe.

Appendix 2 Barra US-E4, GEM4 Earnings Yields and Momentum Strategies

See Figs. 1, 2, 3, and 4

Fig. 1 USA pure earnings yield factor portfolios. Cumulative performance is reported for (a) market-capitalization regression weights, (b) root-cap regression weights, and (c) equal regression weights

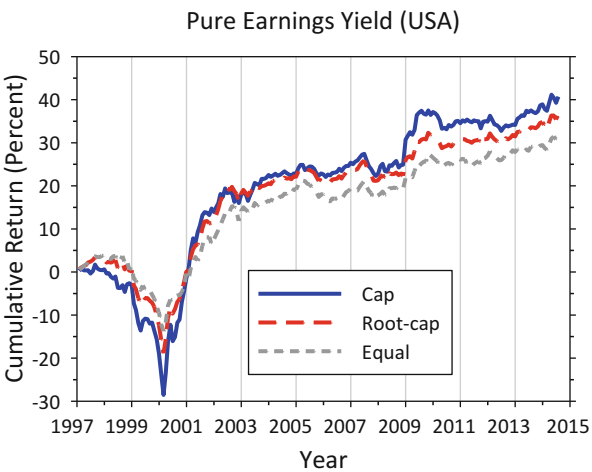


Fig. 2 International pure earnings yield factor portfolios. Cumulative performance is reported for (a) market-capitalization regression weights, (b) root-cap regression weights, and (c) equal regression weights

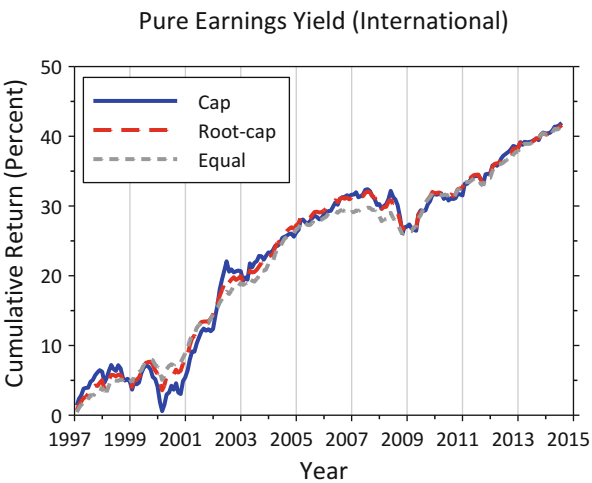


Fig. 3 USA pure momentum factor portfolios. Cumulative performance is reported for (a) market-capitalization regression weights, (b) root-cap regression weights, and (c) equal regression weights

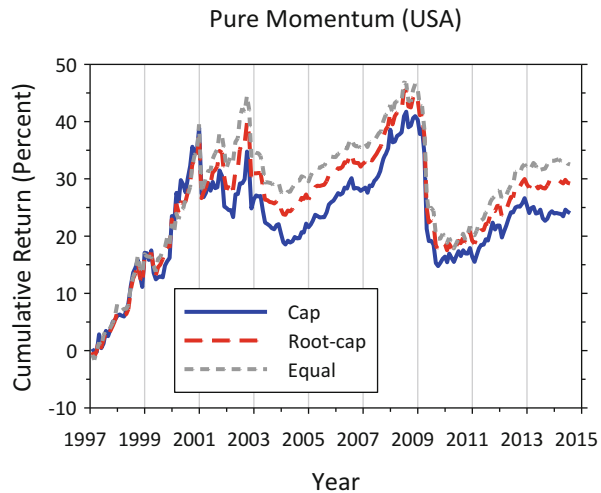
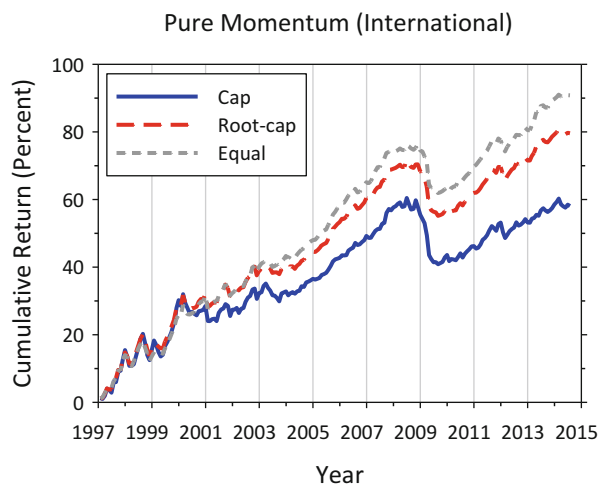


Fig. 4 International pure momentum factor portfolios. Cumulative performance is reported for (a) market-capitalization regression weights, (b) root-cap regression weights, and (c) equal regression weights



References

- Barillas, F., and J. Shanken. 2018. Comparing asset pricing models. *Journal of Finance* 73: 715–755.
- Baumol, W.J. 1963. An expected gain-confidence limit criteria for portfolio selection. *Management Science* 10: 174–182.
- Black, F., M.C. Jensen, and M. Scholes. 1972. The capital asset pricing model: Some empirical tests. In *Studies in the theory of capital markets*, ed. M. Jensen. New York: Praeger.

- Blin, J.M., S. Bender, and J.B. Guerard Jr. 1997. Earnings forecasts, revisions and momentum in the estimation of efficient market-neutral Japanese and U.S. portfolios. In *Research in finance*, ed. A. Chen, vol. 15: 93–114. Greenwich: JAI Press.
- Bloch, M., J.B. Guerard Jr., H.M. Markowitz, P. Todd, and G. Xu. 1993. A comparison of some aspects of the U.S. and Japanese equity markets. *Japan and the World Economy* 5: 3–26.
- Borkovec, M., I. Domowitz, B. Kiernan, and V. Serbin. 2010. Portfolio optimization and the cost of trading. *Journal of Investing* 19: 63–76.
- Bruce, B., and Epstein, C. B. 1994. *The Handbook of Corporate Earnings Analysis*. Chicago: Probus Publishing Company.
- Ceria, S., A. Saxena, and R.A. Stubbs. 2012. Factor alignment problems and quantitative portfolio management. *The Journal of Portfolio Management* 38(2): 29–43.
- Cohen, K.J., and J.A. Pogue. 1967. An empirical evaluation of alternative portfolio-selection models. *Journal of Business* 40: 166–193.
- Conner, G., and R.A. Korajczyk. 1995. The arbitrary theory and multifactor models of asset returns. In *Handbooks in operations research and management science: Finance*, ed. R.A. Jarrow, V. Malsimovic, and W.T. Ziemba. Amsterdam: North-Holland.
- Connor, G., L. Goldberg, and R.A. Korajczyk. 2010. *Portfolio risk analysis*. Princeton: Princeton University Press.
- Dhrymes, P.J., I. Friend, and N.B. Gültekin. 1984. A critical re-examination of the empirical evidence on the arbitrage pricing theory. *Journal of Finance* 39: 323–346.
- Dhrymes, P.J., I. Friend, M.N. Gültekin, and N.B. Gültekin. 1985. New tests of the APT and their implications. *Journal of Finance* 40: 659–674.
- Dimson, E. 1988. *Stock Market Anomalies*. Cambridge: Cambridge University Press.
- Elton, E.J., and M.J. Gruber. 1970. Homogeneous groups and the testing of economic hypothesis. *Journal of Financial and Quantitative Analysis* 5: 581–602.
- . 1973. Estimating the dependence structure of share prices – Implications for portfolio selection. *Journal of Finance* 28: 1203–1232.
- Elton, E.J., M.J. Gruber, and M.W. Padberg. 1978. Simple criteria for optimal portfolio selection: Tracing out the efficient frontier. *Journal of Finance* 33: 296–302.
- . 1979. Simple criteria for optimal portfolio selection: The multi-index case. In *Portfolio theory, 25 years after: Essays in honor of Harry Markowitz*, ed. E.J. Elton and M.J. Gruber. Amsterdam: North-Holland.
- Elton, E.J., M.J. Gruber, S.J. Brown, and W.N. Goetzman. 2007. *Modern portfolio theory and investment analysis*. 7th ed. New York: Wiley.
- Fama, E.F. 1976. *Foundations of finance*. New York: Basic Books.
- Fama, E.F. and K.R. French. 1992. Cross-sectional variation in expected stock returns. *Journal of Finance* 47: 427–465.
- . 1995. Size and the book-to-market factors in earnings and returns. *Journal of Finance* 50: 131–155.
- . 2008. Dissecting anomalies. *Journal of Finance*, 63: 1653–1678.
- Fama, E.F., and J.D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81: 607–636.
- Farrell, J.L., Jr. 1974. Analyzing covariance of returns to determine homogeneous stock groupings. *Journal of Business* 47: 186–207.
- . 1997. *Portfolio management: Theory and applications*. New York: McGraw-Hill/Irwin.
- Grinold, R., and R. Kahn. 1999. *Active portfolio management*. New York: McGraw-Hill/Irwin.
- Guerard, J.B., Jr., and A. Mark. 2003. The optimization of efficient portfolios: The case for a quadratic R and D term. In *Research in finance*, ed. A. Chen. 20: 213–247. Greenwich: JAI Press.
- Guerard, J.B., Jr., R.A. Gillam, H.M. Markowitz, G. Xu, S. Deng, and Z. Wang. 2018. Data mining corrections testing in Chinese stocks. *Interfaces* 48: 108–120.
- Guerard, J.B., Jr., G. Xu, and Z. Wang. 2019. *Portfolio and investment analysis with SAS®: Financial modeling techniques for optimization*. Cary: SAS Press.

- Guerard, J.B., Jr. 2012. Global earnings forecasting efficiency. In *Research in finance*, ed. J. Kenningsser. Bingley: Emerald Publishing Company.
- Guerard, J.B., Jr., and A. Mark. 2020. Earnings forecasts and revisions, price momentum, and fundamental data: Further explorations of financial anomalies. In *Handbook of financial econometrics*, ed. C.F. Lee and J.C. Lee, vol. 2. Singapore: World Scientific Handbook in Financial Economics.
- Guerard, J.B., Jr., and E. Schwartz. 2007. *Quantitative corporate finance*. New York: Kluwer Academic Publishers.
- Guerard, J.B., Jr., and W.T. Ziemba. 2020. *Handbook of applied investment research*. Singapore: World Scientific Publishing.
- Guerard, J.B., Jr., M. Gültekin, and B.K. Stone. 1997. The role of fundamental data and analysts' earnings breadth, forecasts, and revisions in the creation of efficient portfolios. In *Research in finance*, ed. A. Chen, vol. 15. Greenwich: JAI Press.
- Guerard, J.B., Jr., G. Xu, and M.N. Gültekin. 2012. Investing with momentum: The past, present, and future. *Journal of Investing* 21: 68–80.
- Guerard, J.B., Jr., R.T. Rachev, and B. Shao. 2013. Efficient global portfolios: Big data and investment universes. *IBM Journal of Research and Development* 57: 11:1–11:11.
- Guerard, J.B., Jr., H.M. Markowitz, and G. Xu. 2014. The role of effective corporate decisions in the creation of efficient portfolios. *IBM Journal of Research and Development* 58: 1–11.
- . 2015. Earnings forecasting in a global stock selection model and efficient portfolio construction and management. *International Journal of Forecasting* 31: 550–560.
- Guerard, J.B., Jr., H.M. Markowitz, G. Xu, and Z. Wang. 2018. Global portfolio construction with emphasis on conflicting corporate strategies to maximize stockholder wealth. *Annals of Operations Research* 267: 203–219.
- Hirshleifer, D. 2001. Investor psychology and asset pricing. *Journal of Finance* 64: 1533–1597.
- Jacobs, B.I., and K. Levy. 2017. *Equity management: The art and science of modern quantitative investing*. 2nd ed. New York: McGraw-Hill.
- Jensen, M. 1972. *Studies in the theory of capital markets*. New York: Praeger.
- King, B.F. 1966. Market and industry factors in stock price behavior. *Journal of Business* 39: 139–191.
- Lee, C.F., J. Finnerty, J. Lee, A.C. Lee, and D. Wort. 2013. *Security analysis, portfolio management, and financial derivatives*. Singapore: World Scientific Publishing Company.
- Markowitz, H.M. 1959. *Portfolio selection: Efficient diversification of investment*, Cowles foundation monograph no. 16. New York: Wiley.
- . 2000. *Mean-variance analysis in portfolio choice and capital markets*. New Hope: Frank J. Fabozzi Associates.
- Menchero, and Z. Nagy. 2017. Performance of earnings yield and momentum factors in US and International Equity Markets. In *Portfolio construction, measurement, and efficiency: Essays in Honor of Jack Treynor*, ed. J.B. Guerard. New York: Springer.
- Miller, W., G. Xu, and J.B. Guerard Jr. 2014. Portfolio construction and management in the BARRA aegis system: A case study using the USER data. *Journal of Investing* 23: 111–120.
- Mossin, J. 1966. Equilibrium in a capital asset market. *Econometrica* 34: 768–783.
- . 1973. *Theory of financial markets*. Englewood Cliffs: Prentice-Hall.
- Reinganum, M. 1981. The arbitrage pricing theory: Some empirical tests. *Journal of Finance* 36: 313–321.
- Roll, R., and S.A. Ross. 1984. The arbitrage pricing theory approach to strategic portfolio planning. *Financial Analysts Journal* 40 (3): 14–26.
- Rosenberg, B. 1974. Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis* 9: 263–274.
- . 1976. Security appraisal and unsystematic risk in institutional investment. Proceedings of the seminar for research in security prices. University of Chicago.
- Rosenberg, B., and V. Marathe. 1979. Tests of capital asset pricing hypotheses. In *Research in finance*, ed. H. Levy. Greenwich, Ct: JAI Press. 1: 115–224.

- Rosenberg, B., and W. McKibben. 1973. The prediction of systematic and unsystematic risk in common stocks. *Journal of Financial and Quantitative Analysis* 8: 317–333.
- Rosenberg, B., and A. Rudd. 1977. *The yield/beta/residual risk tradeoff*, Working paper no. 66. Research program in finance. Berkeley: Institute of Business and Economic Research, University of California.
- Rosenberg, B., M. Hoaglet, V. Marathe, and W. McKibben. 1975. *Components of covariance in security returns*, Working paper no. 13. Research program in finance. Berkeley: Institute of Business and Economic Research, University of California.
- Ross, S.A. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13: 341–360.
- Ross, S.A., and R. Roll. 1980. An empirical investigation of the arbitrage pricing theory. *Journal of Finance* 35: 1071–1103.
- Rudd, A., and H.K. Clasing. 1982. *Modern portfolio theory: The principles of investment management*. Homewood: Dow Jones-Irwin.
- Rudd, A., and B. Rosenberg. 1979. Realistic portfolio optimization. In *Portfolio theory, 25 years after*, ed. E. Elton and M. Gruber. Amsterdam: North-Holland.
- . 1980. The “Market Model” in investment management. *Journal of Finance* 35: 597–607.
- Saxena, A., and R.A. Stubbs. 2012. An empirical case study of factor alignment using the USER model. *Journal of Investing* 21: 25–44.
- . 2013. Alpha alignment factor: A solution to the underestimation of risk for optimized active portfolios. *The Journal of Risk* 153: 3–37.
- . 2015. Augmented risk models to mitigate factor alignment problems. *Journal of Investment Management* 13 (3): 57–79.
- Sharpe, W.F. 1963. A simplified model for portfolio analysis. *Management Science* 9: 277–293.
- . 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- . 1971. A linear programming approximation for the general portfolio analysis problem. *Journal of Financial and Quantitative Analysis* 6: 1263–1275.
- . 2012. *William F. Sharpe: Selected works*. Singapore: World Scientific Publishing Company.
- Stone, B.K. 1970. *Risk, return, and equilibrium: A general single-period theory of asset selection and capital market equilibrium*. Cambridge, MA: MIT Press.
- . 1973. A linear programming formulation of the general portfolio selection problem. *Journal of Financial and Quantitative Analysis* 8: 621–636.
- . 1974. Systematic interest-rate risk in a two-index model of returns. *Journal of Financial and Quantitative Analysis* 9: 709–721.
- Ziemba, W.T. 1993. *Invest Japan*. Chicago: Probus Publishing Co.



Short Selling Activity and Effects on Financial Markets and Corporate Decisions

98

Xu Guo and Chunchi Wu

Abstract

We review the current state of research on short sales, including the information content in short interest, information advantages of short sellers, and firms' reactions to the short-selling threat. We also discuss the data and metrics commonly used in short selling studies and summarize important empirical findings. Finally, we suggest several future directions to improve our understanding of the role of short sellers in capital market development and in influencing corporate decisions.

Keywords

Short interest · Short selling · Return predictability · Corporate decisions · Financial reporting

JEL Classification

G12 · G13

98.1 Introduction

Short selling activities are prevalent in the US financial markets, and there has been a substantial increase in shares held short by investors in the past several decades. A vast literature has been devoted to understanding the effects of short selling activity. Despite the enormous amounts of studies, the role of short sellers in financial

X. Guo

College of Economics, Shenzhen University, Shenzhen, China

e-mail: xuguo@szu.edu.cn

C. Wu (✉)

School of Management, University at Buffalo, Buffalo, NY, USA

e-mail: chunchiwu@buffalo.edu

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,

https://doi.org/10.1007/978-3-030-91231-4_100

2313

markets remains controversial. On the one hand, they are blamed for manipulating and depressing share values, and increasing market volatility and instability (Allen and Gale 1992; Brunnermeier and Oehmke 2014; Goldstein and Guembel 2008; Ljungqvist and Qian 2016). On the other, they are credited with improving financial markets' information efficiency and strengthening the price discovery process in financial markets and corporate governance (Boehmer et al. 2008; Diether et al. 2009a; Boehmer and Wu 2013; Curtis and Fargher 2014). Because of its important role, issues related to short sales have continued to attract considerable attention from academics and policy makers.

In a short sale, investors or short sellers go against the common trading rule by selling a security with the expectation of buying it back after a price drop, realizing a profit from the loss in the share price. Typically, short sellers do not own the shares before the transaction, but borrow (equity loan) them from another investor or institution. The risk to the short seller is that the security's price could increase and he or she may have to buy it back at a higher price. In addition, short sellers are exposed to the recall risk, which refers to the situation when the borrowed stock is recalled by the lender unexpectedly and a short seller is forced to close his position.

Miller (1977) suggests that dispersion of investor opinions in the presence of short sale constraints leads to stock overvaluation, as short sale constraints impede the adjustment of stock prices to negative private information. Short sale constraints include various costs and risks of shorting, as well as legal and institutional restrictions. Generally, it is easy and cheap to borrow most large cap stocks, but it can be difficult to borrow small firm stocks that have smaller amounts of share supply, have low institutional ownership, or are in high demand from borrowers. Aside from the problems in the stock lending market, there are a variety of other short sale constraints. US equity markets are not set up to make shorting easy. Regulators, the Federal Reserve, various stock exchanges, and individual brokerage firms can mechanically impede short selling. For example, the uptick rule on the NYSE, established by the Securities and Exchange Commission (SEC), requires short sales to be conducted at a higher price than the previous trade or at an uptick. For NASDAQ stocks, the bid-price rule requires that a short sale must be executed above the previous inside bid.¹ In addition to regulations, short sellers face hostility from society at large, especially during the financial crisis.

There is an extensive literature exploring the relationship between short interest (as a proxy for short selling activity) and future stock abnormal returns. Earlier studies provide mixed evidence, but the recent literature tends to reach a consensus that short sellers are indeed informed traders. There is also substantial evidence that short selling activity influences corporate and accounting decisions. Short sellers play a disciplining role by exposing corporate fraud and force managers to take corrective actions.

¹Shorting restrictions can date back to the 1930s. For example, the 1931 downtick shorting prohibition, the 1932 rule that requiring written authorization to lend shares, and the strict uptick rule in 1938. See Jones (2012) for more details.

A large literature focuses on the information advantages of short sellers, which confirms their superior ability to collect and process public information and acquire private information. The sources for public information are financial statements (Dechow et al. 2001; Hirshleifer et al. 2011; Park 2017) or media coverage (Hobbs et al. 2012). There is considerable evidence that short sellers can successfully predict a series of bad corporate events such as analysts' downgrades, earnings restatements, financial fraud, insider sales, etc., suggesting that private information plays an important role in the actions of short sellers.

Many studies explore firms' behaviors in the presence of a short selling threat. To mitigate the adverse effect of price pressure caused by elevated short interest, managers attempt to improve their financial reporting quality (Fang et al. 2016), increase cash dividends (Chen et al. 2019), reduce investment and equity issues (Grullon et al. 2015) while simultaneously improving investment efficiency (Deng et al. 2020).

In this chapter, we provide a comprehensive review on the functions of short sales and the short-selling process, short sales activities and regulations, theoretical explanations for short sales, and important empirical evidence. In addition, we review the current state of short sale research and discuss possible future directions.

The remainder of this chapter is organized as follows. In Sect. 98.2, we discuss the functions and processes of short sales. In addition, we describe the rationales for short selling as well as various short sale constraints. In Sect. 98.3, we present theories of short selling, and in Sect. 98.4, we review various functions of short sales short selling. In Sect. 98.5, we report important empirical evidence in more detail and discusses the data/metrics used in short-selling studies. Finally, we offer suggestions for future research in Sect. 98.6 and conclude in Sect. 98.7.²

98.2 The Short-Selling Process, Institutional Structure, and Activities

98.2.1 Why Going Short?

A number of studies have attempted to explain why going short in the first place. The literature has suggested that underlying fundamentals, arbitrage, hedging, speculation, and tax-related considerations are important factors behind the motivation of short sellers. Dechow et al. (2001) find that firms with low ratios of fundamentals (e.g., earnings and book values) to market values have high short interest. Geczy et al. (2002) show that short selling can be driven by stylized patterns and behavior in stock prices, such as the size, book-to-market, and momentum effects. On the other

²To our best knowledge, Reed (2013) and Jiang et al. (2020a) also give a review of the short-sale literature.

hand, Lamont and Stein (2004) show that short interest does not seem to be related to market-wide valuation trends.

Brent et al. (1990) suggest that short selling can be triggered by factors such as deferring capital gains taxes, hedging needs, exploiting the overpriced stocks, or speculation. The tax motivation refers to the situation that when going short in the same stock that is held long, an investor can lock in a profit, but delay the recognition of a capital gain, which is particularly important if the investor will be taxed at a lower rate in the subsequent tax period. Consequently, based on the tax motivation, one should observe an increase in short interest primarily at the end of the tax year (Dyl 1978).

Arbitrage and hedging are also important in explaining short interest as going short is an integral part of arbitrage strategies. For example, investors can engage in shorting the acquiring firm's stock in a merger situation while simultaneously going long in the target firm's stock (Baker and Savaşoglu 2002). Short sellers can also sell securities short to arbitrage a price differential between the stock and the corresponding convertible bond.

Investors can also speculate through shorting and expect to profit from a decline in the security's price. Short sellers have beliefs that deviate from consensus beliefs impounded in the existing prices, which is consistent with the view that short interest is a measure of heterogeneous investor opinion. The existing literature supports the argument that short sellers help to facilitate the price discovery process, which is attributed to their extraordinary skills in exploiting both public and private information.

98.2.2 The Security Lending Process

To do short selling, a potential short seller must find a party who is willing to lend the shares. In the USA, most exchange procedures require that the short seller delivers shares to the buyer on the third day after the transaction ($t + 3$). The proceeds from the short sale are deposited with the lender of the stock. For stocks, the lender typically requires 102% of the loan's value in collateral. The value of the security loan is marked to market daily. An increase in the stock price will result in the lender requiring additional collateral for the loan. Conversely, a decrease in the stock price will result in the lender returning some of the collateral to the borrower. All collateral must return to the borrower upon the return of shares to the lender. The process of delivery and payment is referred to as settlement or clearing. While a stock is on loan, the lender invests and receives interest on the collateral. Part of this earned interest goes to the borrower in the form of a negotiated rebate rate.

98.2.3 Short-Selling Activities in the Market

The equity lending market is an over-the-counter market. Detailed data on the aggregate short-selling activities are scarce. Aggregating a monthly firm-level

short interest data into a short interest index, Rapach et al. (2016) show there is a strong upward trend in the short selling activity in the USA. They argue that the expanded equity lending market as well as the significant increase in the amount of capital devoted to short arbitrage could be the explanation for their finding. Indeed, short selling is a common phenomenon in stock markets and there are many short sale orders executed in margin accounts on any given day. Using a randomly selected sample of common stocks from the CRSP, Asquith et al. (2010) report that short sales account for 29.7% of all stock trades and 27.9% of the total trading volume. Diether et al. (2009b) report similar findings with short sales representing 24% of NYSE and 31% of NASDAQ volume. Short sellers also trade in newly issued stocks. Edwards and Hanley (2010) report that short selling occurs on the offer day in 99.5% of the IPOs they studied.

98.2.4 Short Sale Constraints

Short selling can be expensive (high borrow cost), difficult (hard to find lender), and risky (potential loss). Impediments to short selling include the expense and difficulty of borrowing stocks, legal and institutional restrictions, regulation rules and trading mechanisms, and the risk that the short position may be involuntarily closed due to a recall by the lender. Yet, there are other forms of short sale constraints. Firms targeted by short sellers can take a variety of actions to impede the short selling of their stocks. For example, firms can accuse short sellers of illegal activities and hire private investigators to probe them or even request authorities to investigate their activities. Lamont (2012) studies legal battles between short sellers and firms. Using a sample of 266 firms who threaten, take action against, or accuse short sellers of illegal activities or false statements, the author shows that the sample firms earn negative abnormal returns in the subsequent year as these actions create short sale constraints that lead to overpriced stocks.

Past studies report mixed findings in relation to the benefits of short selling constraints, especially those regulatory restrictions of short selling. Brunnermeier and Oehmke (2014) show temporary and targeted short sale restrictions can save vulnerable stocks. On the other hand, a number of studies suggest that short selling constraints distort the market efficiency by slowing down the impounding of negative private information (Choy and Zhang 2019; Ho 1996; Hasan et al. 2015; Shyu et al. 2018; Cheng et al. 2012), reduce hedging effectiveness (Brunnermeier and Oehmke 2014), and contribute to IPO underpricing (Boulton et al. 2020).

Generally, for stocks with tight short sale constraints, the overpricing can be very large. Consistent with this view, Asquith et al. (2005) investigate a group of constrained stocks with a strong demand to sell short and a limited supply of shares to borrow, that is, the short sale constraints are binding for these stocks. They find the constrained stocks underperform by more than 2% per month. However, they point out short interest and institutional ownership could not be short sale constraints for the majority of stocks.

Short sale constraints are used to explain the idiosyncratic volatility puzzle: stocks with high idiosyncratic volatility (IVOL) are associated with low future returns. Stambaugh et al. (2015) argue the IVOL-return relation is negative among overpriced stocks but positive among underpriced stocks, but due to the arbitrage asymmetry caused by short sale constraints, the negative relation among overpriced stocks dominates, resulting in an overall negative relation. In a separate vein, Jiang et al. (2014) find investor overconfidence with binding short sale constraints can explain the negative IVOL-return relation, and they use three natural experiments to support their argument.

98.2.5 Short-Selling Costs and Rebate Rates

It is well documented in the literature that short positions can be expensive or impossible and can be involuntarily terminated. An investor must generally borrow shares they shorted, and then deliver them to the buyer. Thus, the cost, availability, and dependability issues related to shorting generally result from the pricing (loan fees), availability (loan supply), and termination of equity loans.

An equity loan is a temporary swap of ownership. The equity lender transfers legal ownership of a block of shares to the borrower, who simultaneously transfers collateral (usually cash). Upon the loan origination, the parties negotiate a rebate rate, which is the amount of interest on the collateral that the lender will rebate to the borrower. Thus, a high rebate rate indicates a low borrowing cost, *ceteris paribus*. Jones and Lamont (2002) argue that the rebate rate is probably the best proxy for measuring the magnitude of short sale constraints. However, because the stock lending market is extremely opaque, rebate rates are never transparent.

Using the proprietary data from a large financial institution, D'avolio (2002) finds overall, it is easy to borrow stocks with an average loan fee of about 25 basis points per annum. They also find less than 10% of stocks are expensive to borrow (loan fees above 1% per annum) and the value-weighted mean loan fees are only 17 basis points for the remaining 90% of stocks. These results suggest that stocks are generally inexpensive to borrow. Geczy et al. (2002) study another proprietary equity loan data and report similar findings for the low cost of short selling.

Chague et al. (2017) study the Brazilian equity lending market and provide evidence that the lending market is relationship-based: well-connected borrowers face low search costs and pay lower loan fees, which confirm the prediction made in Duffie et al. (2002). Engelberg et al. (2018) use the variance of loan fees as a proxy for short selling risk and find that stocks with higher short selling risk are associated with lower future returns, poorer price efficiency (measured by price delay as in Hou and Moskowitz (2005)), and less short selling. Evans et al. (2009) and Lin and Lu (2016) show empirically, and Atmaz and Basak (2019) demonstrate theoretically a higher shorting cost leads to higher bid-ask spreads for typical options, and higher apparent put-call parity violations.

98.3 Theory of Short Sales

Miller (1977) first shows that the presence of short sale constraints impedes the adjustment of stock prices to negative information, which results in higher stock prices when investor beliefs are heterogeneous. Along a similar line, Hong and Stein (2003) argue short sale constraints sideline bearish investors to trade and lead to elevated stock prices if investors are of different opinions. When the market is falling, the private information of bearish investors accumulates and heightens the likelihood of a market crash. Hong et al. (2006) link short sale constraints, heterogeneous beliefs, limited asset float (tradable shares) to speculative bubbles.

In a separate vein, Jarrow (1980) shows that short sale constraints can cause stock prices to rise or fall if investors have heterogeneous expectations (beliefs) but disagree about the covariance matrix of future prices. Short sale constraints will inflate risky asset prices only under the condition of homogeneous beliefs for the covariance matrix of future prices. Diamond and Verrecchia (1987) find short sale constraints eliminate informative trades but do not bias stock prices upward if investors are rational and fully aware of the effects of short sale constraints.

Duffie et al. (2002) model the structure of the equity lending market in which search cost is a key variable. The existence of search cost lowers the frequency at which borrowers can find lenders, resulting in lenders acting as monopolists and charge higher loan fees. Moreover, if lendable securities are difficult to locate, then the price of the security is initially elevated, and the price could be even higher than the most optimistic buyer's valuation due to the prospect of lending fees. As time passes, short interest increases and unfilled shorts are reduced, so that each share is expected to be lent at a smaller number of future times, which depresses the lending fee and causes the future price to decline. The Duffie et al. (2002) model offers an explanation for the dynamic price behavior of stocks sold short in a market where opacity increases the frictions of short sellers.

Atmaz and Basak (2019) provide a dynamic analysis of option prices in the presence of costly short selling and option market makers. They find that call and put bid-ask spreads increase with the shorting fee for typical options because an increase in the shorting fee not only raises short-selling costs but also partially increases the benefit of holding a share long, leading to higher bid-ask spreads for typical options. They further show that put bid and ask prices as well as the put option implied volatilities, increase with the shorting fee as a higher shorting fee increases both the market makers' cost of and the proceeds from the hedge portfolios as a put seller and buyer, respectively. As implied stock prices decrease in the shorting fee, there is a larger deviation from the underlying stock prices, leading to higher apparent put-call parity violations, since the implied stock is equivalent to the stock but without the short-selling costs and lending benefits.

Moreover, Atmaz and Basak (2019) show that the effects of short-selling costs on option bid-ask spreads are more pronounced for relatively illiquid options with lower trading activity. The reason is that market makers are more likely to hedge the relatively illiquid options via hedge portfolios, through which short-selling costs affect option prices directly. In an extended analysis, they show that higher stock

short-selling costs lead to lower corporate bond prices or higher yields. In addition, stock short-selling fees increase bid-ask spreads for corporate bonds.

98.4 Literature Review

98.4.1 Short Sale Supply

Beneish et al. (2015) explore the determinants and consequences of the short sale supply. They find that the lendable supply is sensitive to borrowing costs and firm characteristics. Special (hard-to-borrow) stocks with the lowest supply have the strongest negative returns. In addition, they find that the short side returns of nine well-known market anomalies are due to those constrained, hard-to-borrow special stocks. For stocks that are not special, the short side returns are less significant. Finally, if short sellers' costs are considered, the profitability of the short side is severely limited. However, in the equity lending market, lendable supply, demand for borrowing, the price for borrowing, quantity borrowed, and short interest are closely related (Reed 2015), and lendable supply should not be considered as a distinct element. This endogeneity problem is explained in greater detail in Reed (2015). He considers a potentially new take and suggests that one should consider short selling demand as one of the only exogenously determined variables in the equilibrium relationship.

Several papers use natural experiments to address the endogeneity issue in short sale tests. Thornock (2013) examines the effects of dividend taxation between the lender and the short seller. Lenders often reduce loan supply around the dividend record date since after-tax dividends usually exceed the substitute payments reimbursed from short sellers. The tax-driven exogenous shock allows the paper to determine the effect of loan supply on loan prices and short selling. Using the same experimental setup, Blocher et al. (2013) find that when supply is reduced around dividend record dates, prices of hard-to-borrow stocks increase 1.1% while prices of easy-to-borrow stocks are unaffected. Moreover, they document the hard-to-borrow stocks are associated with lower 3-month returns.

98.4.2 Cross-Sectional Return Predictability

The short interest in a stock is often viewed as a measure of the heterogeneity of investor opinion. A number of studies have investigated the information content in equity short interest, and the negative relationship between the short interest levels and future stock returns is well-established (Desai et al. 2002; Nagel 2005; Diether et al. 2009b; Boehmer et al. 2010; Takahashi 2010; Pownall and Simko 2005). For example, Desai et al. (2002) examine the relationship between the short interest levels and stock returns of NASDAQ firms. They find firms with a high level of short interest experience negative and significant abnormal returns. Boehmer et al. (2010), on the other hand, find the absence of short selling is a positive signal about future

returns. Moreover, these positive future returns are larger in magnitude than the negative returns on heavily shorted stocks. The key question is how we should interpret the fact that short sellers can successfully time their trades when the stocks are overvalued.

Diether et al. (2009b) provide several possible explanations. The first one falls into the informed trading story that short sellers are either corporate insiders or traders who can have access in advance to material nonpublic information. The second explanation is that short sellers engage in short selling to provide liquidity. When there is a significant and temporary buy-order imbalance in the market, short sellers step in and trade. As the buying pressure extenuates or even vanishes, prices will go down and short sellers make a profit. The third explanation, supported by the model of Diamond and Verrecchia (1987) that argue short sales by liquidity traders are less likely due to the costly nature of short selling, recognizes that short sellers are sophisticated traders. As such, short sellers can successfully locate and exploit overpriced stocks, when the overpricing could be triggered by market frictions or behavioral biases. The last explanation is that short sellers play a role in providing additional risk-bearing capacity in periods of elevated uncertainty.

Short sellers also present skills in exploiting industry information. Huszár et al. (2017) find most-shorted stocks in industries with the highest aggregate shorted values substantially underperform other most-shorted stocks in less shorted industries. Short sellers specifically target complex or informationally diverse industries to maximize their profits.

98.4.3 Time Series Return Predictability (Aggregate Short Interest)

Another strand of research investigates the relationship between market-level short sale activities and market returns. An early study by Seneca (1967) finds that high aggregate short interest amount is associated with a low S&P 500 index level. Using daily short interest data, Lynch et al. (2014) also document that an increase in aggregate shorting is associated with a decrease in market excess returns over the following 5–20 days. Lastly, the detrended aggregate short interest, denoted as short interest index (SII) in Rapach et al. (2016), is a powerful predictor of aggregate stock returns. High detrended short interest predicts low market returns.

98.4.4 Short Interest and Corporate Events/Firm Performance

The information advantages of short sellers arise from their abilities to predict a variety of bad corporate events from changes in analysts' earnings estimates and recommendations (Christophe et al. 2010; Meng et al. 2017) to negative earnings surprises (Christophe et al. 2004), earnings restatements (Desai et al. 2006), financial frauds (Karpoff and Lou 2010), insider sales (Chakrabarty and Shkilko 2013; Khan and Lu 2013), private placements (Berkman et al. 2017), asset write-downs (Liu

et al. 2012), negative operating performance (Deshmukh et al. 2015; Akbas et al. 2017), and accounting irregularities (Dechow et al. 1996).

Christophe et al. (2010) study short selling prior to the release of analyst downgrades and document abnormal levels of short selling in the 3 days before downgrades are publicly announced. Using the data from China, Meng et al. (2017) find evidence of abnormal short selling in the days before downgrade announcements that are made by star analysts.

Christophe et al. (2004) study short selling activities 5 days prior to earnings announcements. For a subset of firms with negative earnings surprises, their pre-announcement short selling is significantly higher, which suggests the existence of informed traders. Christensen et al. (2014) find short selling is significantly high around earnings announcements that contain pro forma earnings disclosures, suggesting short seller view pro forma earnings disclosures as a strong signal of overvaluation. Indeed, firms that make pro forma earnings disclosures and have high short interest experience negative abnormal returns subsequently.³

Desai et al. (2006) examine the behavior of short sellers around the time when firms publicly acknowledge material errors in their previously reported financials (earnings restatements) and link short selling to questionable financial reporting. The results of their event study show that short sellers can identify the restatements before their public disclosure, and they build up short positions in the restating firms in the months leading up to the restatement announcement.⁴

Karpoff and Lou (2010) find that short sellers can detect financial misconducts. In anticipating the revelations of such misrepresentations, short sellers accumulate their short positions during the 19-month period before the revelations. Chakrabarty and Shkilko (2013) document that short sellers can detect insider trading. Specifically, short selling increases before insider sales are publicly reported and often before insiders finish selling. The superior timing ability of short sellers stems from the monitoring of order flow and price-relevant information received from brokerages that execute insider sales. In addition, Khan and Lu (2013) find evidence that short sellers front-run large insider sales.

Berkman et al. (2017) show there is a significant increase in the short interest surrounding the preannouncement of stock private placements, suggesting short sellers receive information from the insiders. Financial firms have experienced large asset write-downs during the 2007–2008 subprime mortgage crisis. Liu et al. (2012) find short sellers accumulate short positions prior to write-down announcements, suggesting the existence of information advantages of short sellers.

Short sellers also possess superior skills in exploiting public information, such as news releases, past stock prices, financial statements, and media information.

³When standard GAAP earnings are released, companies sometimes announce an adjusted earning that excludes one or more income statement line items from net income, which is known as pro forma earnings. The exclusion of some transitory items from GAAP income will overstate operating results and even misleads investors.

⁴However, Drake et al. (2015) find that short sellers do not appear to anticipate restatement announcement, but take actions when restatements are announced.

Engelberg et al. (2012) show the well-documented negative relation between short interest and future returns is twice as large on news days and four times as large as on days with negative news. Lee and Piqueira (2017) find short selling is negatively associated with the nearness of the price to the 52-week high. Dechow et al. (2001) show that short sellers target stocks with low ratios of fundamentals to market values that are likely to have systematically lower future returns. Accrual anomaly suggests that firms with high operating accruals will have low future abnormal returns (Sloan 1996). Thus, short sellers are inclined to target firms with high accruals. Hirshleifer et al. (2011) find a positive association between short selling and accruals during 1988–2019. In addition, short sellers also target firms with more earnings management activities (Park 2017) or firms that are about to disclose first-time internal control material weaknesses (ICMWs) under Section 404 of the Sarbanes-Oxley Act.

Drake et al. (2011) show short sellers interpret 11 variables properly in accordance with how these information variables affect future stock returns, whereas analysts tend to positively recommend stocks with high growth, high accruals, and low book-to-market ratios, which have a negative relationship with future returns.⁵ There is evidence that short sellers pay attention to media information. Hobbs et al. (2012) find a positive relationship between short selling and *Mad Money* buy recommendations, which are assembled from Jim Cramer's television show. The recommended stocks experience a positive abnormal return in the short run followed by a quick reversal; short sellers thus step in and seize this opportunity to exploit the short-term overpricing arising from the behavioral biases.

Despite the information advantages that are possessed by short sellers, which allow them to take actions before various adverse corporate events become public, short sellers also trade after the information is released. For example, Blau et al. (2013) show short selling activities increase significantly on and shortly after the announcement of an auditor change. Berkman and McKenzie (2012) find short sellers take short positions immediately after earnings announcement. These findings suggest that short sellers may have multiple information signals or they can update their information conditional on the new information released.

98.4.5 Consequences of Short Selling

In the presence of short selling threats, firms will likely change their investment activities, financing decisions, tax avoidance, and other corporate policies to cope with it. For example, Chen et al. (2019) find pilot firms that are more vulnerable to the short-selling threat will increase their cash dividends during the Reg SHO program,

⁵The 11 variables are earnings surprise, total accruals, capital expenditures, market-value-of-equity, earnings-to-price ratio, book-to-market ratio, stock turnover, sales growth, forecasted long-term growth, earnings forecast revision, and price momentum, which are all constructed from publicly available information.

which conveys a signal to the market that stocks are undervalued. Also under the Reg SHO program, Grullon et al. (2015) show that the constrained firms in the pilot program reduce investment and equity issues when there is a big increase in short selling activity. Deng et al. (2020) find pilot firms improve investment efficiency, which results in a decrease in stock price crash risk. Chen et al. (2020) report pilot firms issue more long-run good news forecasts to discourage short sellers. Similarly, Li and Zhang (2015) discover pilot firms will reduce the precision of bad-news forecasts and the readability of bad-news annual reports in response to short selling threats. Fang et al. (2016) find pilot firms are less likely to engage in earnings management activities. De Angelis et al. (2017) show pilot firms tend to grant more stock options to the managers to alleviate the adverse effects of the bear raids by incentivizing the managers. Last but not least, Gong (2020) document firms choose conservative financial policies and reduce corporate leverage when facing short-selling threat.

Tax avoidance is a bad signal to the outsiders, and firms are thus less willing to engage in corporate tax avoidance strategies, hoping to impede the short selling (Guo et al. 2018). However, their findings fail to reconcile the fact that tax avoidance is also beneficial for shareholders (Edwards et al. 2016).

Corporate social responsibility (CSR) has become increasingly important in recent years. Hou et al. (2019) find firms are more likely to engage in corporate philanthropic activities in response to an increase in short selling activities through investigating data from China.

98.4.6 The Role of Short Interest in Other Asset Markets

Short interest also appears to play an important role in the corporate bond market. Specifically, Kecskés et al. (2013) find that firms with high short interest have lower credit ratings, and the bonds are more likely to have their ratings downgraded. Moreover, firms with highly shorted stocks are associated with a higher cost of debt. Henry et al. (2015) also show that short sellers start to build up their positions several months ahead of a bond downgrade. Moreover, they provide empirical evidence of informed trading as short interest is higher for unanticipated downgrades. In addition to rating changes, Shi et al. (2017) study short selling and crediting rating agency (CRA) announcements, which include rating changes and credit watch placements. They find that short selling predicts negative credit watches.

Last but not least, equity short interest is a powerful predictor in the CDS market. Griffin et al. (2016) document a negative relation between short interest and future returns on credit default swaps.

98.4.7 Short Sale Ban in the History

Despite the controversial opinions against short sellers, for the most part, they are considered to be the “good guys.” A large body of literature has shown that short sellers are sophisticated (informed) traders that help markets correct short-term

deviations of stock prices from the fundamental value and contribute to efficient stock prices. However, when share prices fall sharply, they tend to be the scapegoats. In September 2008, the US Securities and Exchange Commission (SEC) surprised the investment community by adopting an emergency order that temporarily banned most short sales in nearly 1,000 financial stocks.⁶ Grundy et al. (2012) find the short sale ban is associated with a decline in trade in put options on the banned stock, suggesting option strategies are not effective substitutes for short sales during a short sale ban.

The last time shorting was banned in the United States was in September 1931, when the NYSE banned all short sales in the wake of England's announcement that it was abandoning the gold standard. Boehmer et al. (2013) study this unexpected event and find the market quality was severely compromised. The ban had substantial effects on larger stocks, and those stocks suffered severe degradation in market quality (i.e., effective spreads, price impacts, and realized spreads).

The short selling bans are not solely implemented in the United States. Beber and Pagano (2013) study the short sale bans that 30 countries imposed during the period of the global financial crisis from 1 January 2008 to 23 June 2009. They find short sale bans or regulatory constraints on short selling are harmful to liquidity and slows the price discovery process. Frino et al. (2011) study the 2008 short selling bans on 14 equity markets and document a similar negative impact.

98.5 Data and Empirical Evidence

In this section, we first describe several databases that are commonly used in short selling studies. Following this, we discuss important empirical evidence.

98.5.1 Reg SHO Data – NYSE Short Sales (TAQ)

On 23 June 2004, the SEC adopted Regulation SHO to establish uniform locate-and-delivery standards, created uniform marking requirements for sales of all equity securities, and temporarily suspended the price tests for a set of pilot securities.⁷ The SEC mandated that all self-regulatory organizations make tick data on short sales publicly available starting from 2 January 2005. Short sales reported to the NYSE for NYSE-listed and trades securities can be accessed from the TAQ tape of the WRDS research database. The sample period is from 2 January 2005 to 6 June 2007. Only trades that have a short sale component are included. The data contain ticker, price,

⁶See Boehmer et al. (2013) for more details and the timeline of events.

⁷Sell orders are marked as “long,” “short” “short exempt” where “long” refers to regular sell, “short” refers to short sale, and “short exempt” refers to short sale exempted from the uptick rule regulation. More information on Reg SHO can be found at <https://www.sec.gov/investor/pubs/regsho.htm>

short size, time, listing market, and short type (exempt or nonexempt from the short sale rules). Short trade prints that contain both exempt and nonexempt components are indicated on separate records.

98.5.2 Supplemental Short Interest File (Compustat)

Compustat provides supplement short interest files from 1973. All member firms of NYSE, AMEX, and NASDAQ are required to report the aggregate number of shares sold short over all accounts as of the middle of each month. The Compustat Supplemental Short Interest file contains data for NYSE and AMEX firms beginning in 1973 and for NASDAQ beginning in 2003. Note that raw short interest numbers are reported twice a month from 2007, one in the middle and the other at the end of the month.

98.5.3 Markit Data Explorers (DXL) Equity Lending Data

DXL (Data Explorers was acquired by Markit in 2012) is a research firm that collects data from the securities lending desks at more than 100 of the world's leading institutional traders, which collectively represent the largest pool of loanable equity inventory in the world. Lending data at the stock level are available from May 2002. DXL provides the data at monthly frequency in 2002 and 2003, at weekly frequency from July 2004, and at daily frequency after 2006. Studies that exploit this database include Beneish et al. (2015), Chang et al. (2019), Massa et al. (2015), Saffi and Sigurdsson (2011), etc.

98.5.4 WRDS European Short Data

Following the global financial crisis of 2008, European Securities and Markets Authority (ESMA) launched policy EU N236/2012 on short selling and certain aspects of credit default swaps. The purpose of this EU-wide reporting regulation is to increase market stability by reducing the opaqueness of market activities through mandatory daily reporting by institutional investors. Significant net short positions in shares must be reported to the relevant competent authorities when they are at least equal to 0.2% of company issued share capital and every 0.1% above that, and disclosed to the public when they are at least equal to 0.5% of company issued share capital and every 0.1% above that. The coverage includes Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Poland, Spain, Sweden, and the UK. As each of the participating EU markets has its own site for data downloading and there is not a centralized database that consolidates this relatively new data, WRDS scrape the participating markets reporting website to collect the data on a quarterly basis, which starts from 1 January 2010. The data contain variables such as

country, issuer (the underlying stock), position holder, and its net position (percentage of shares shorted by the position holder as a fraction of issuer's total shares outstanding).

98.5.5 Short Selling Metrics

Two common metrics are used to track the short selling activity on a stock.

- (a) Short interest ratio (SIR): Also known as the short interest level, measures the ratio of shares that are currently being shorted compared to the total number of shares outstanding (see Senchack and Starks 1993; Desai et al. 2002; Boehmer et al. 2010; Shkilko et al. 2012). A high SIR is a negative signal, which is associated with future negative abnormal returns.⁸
- (b) Short interest to trading volume ratio: Also known as the days-to-cover ratio, defined by the total shares that are held short divided by the average daily trading volume (see Diether et al. 2009b). A high days-to-cover ratio is also a bearish indicator.⁹

Short selling activities have increased over time, which could be attributed to many market-wide institutional factors, such as the ease of short selling as a result of continued development of the equity lending market, an increase in the number of hedge funds in existence, and an increase in the supply of loanable shares. In light of this phenomenon, many studies try to detrend the short interest, as the trend in short interest would obscure the true information content buried under the short interest. For example, Rapach et al. (2016) construct an aggregate short interest index and suggest that it is the strongest predictor of aggregate stock returns. They first compute the monthly equal-weighted mean of all stock-level short interest ratio (*EWSI*), which is further detrended using the following regression:

$$\log(EWSI_t) = a + bTREND_t + u_t \quad (1)$$

The standardized fitted residual from the above estimation is the aggregate short interest index, which can be interpreted as a measure of market pessimism.

On the other hand, Guo and Wu (2019) take a simple approach by taking the first difference in SIR and find that the change in the short interest ratio is more

⁸Purnanandam and Seyhun (2018) use a standardized measure of SIR, which reflects the deviation from the average level of shorting of a stock after accounting for its variance.

⁹Traders could get the idea of potential future buying pressure from this ratio. Short sellers usually want to buy the shares back at the lowest price possible. However, it will drive prices higher as short sellers buy to cover. The longer the buyback process takes, as referenced by the 'days to cover' metric, the longer the price rally may continue based solely on the need of short sellers to close their positions. Additionally, a high 'days to cover' ratio often signal a potential short squeeze, a situation that stock price increases rapidly due to short sellers covering their positions.

informative. They also document that the predictability of short interest for future returns is concentrated in worst-rated stocks, similar to the momentum anomaly (Avramov et al. 2007).

There are other short selling metrics used in the literature; for example, Bao et al. (2019) construct the residual short interest, which are the residuals from the following regression:

$$SIR_{i,t} = \alpha + \beta_1 IO_{i,t} + \beta_1 CONVERT_{i,t} + \beta_3 TREND_t + \epsilon_{i,t} \quad (2)$$

where *SIR* is short interest, *IO* is the institutional ownership, which is a proxy for the supply of loanable shares, *CONVERT* is a dummy variable that indicates whether firm *i* has convertible bonds or convertible preferred stocks outstanding, which can control the hedging-based shorting that is unrelated to negative private information. Finally, the time trend in short interest is removed. The regression residual is then used as a measure of negative private information.

Lamont and Stein (2004) study the aggregate short interest during the dot-com bubble period and the aggregate short interest is defined as the scaled value-weighted short-interest ratio (100 times the market value of shares sold short, divided by the value of shares outstanding). They find total short interest moves in a countercyclical pattern. Unlike the contrarian nature of the stock-level short interest, aggregate short interest displays extrapolative behavior: fewer investors are willing to bet on the market going down after a period in which it has been rising, such as the dot-com bubble period. This puzzle could be explained by the open-end nature of professional money management: fund managers are forced to reduce their aggregate short positions due to the redemptions from their clients, who are unwilling to lose money when the market is rising.

98.5.6 Empirical Evidence

Desai et al. (2002) find firms with a high level of short interest experience negative and significant abnormal returns when examining the relationship between the short interest level and stock returns of Nasdaq firms. The monthly short interest data are obtained from the Nasdaq. Using a calendar-time portfolio approach, at the beginning of each month, they form equal-weighted portfolios of stocks based on four thresholds of short interest in the previous month. For the 2.5% portfolio (i.e., stocks have at least 2.5% short interest in the previous month), the risk-adjusted return is -0.76% per month with a *t*-value of -3.34 . The risk-adjusted return decreases monotonically if higher thresholds of short interest level are used in constructing portfolios. For example, limiting the sample to stocks with at least 10% short interest level, the risk-adjusted return is -1.13% per month. Moreover, heavily shorted firms are more likely to incur liquidations or forced delisting than their size-, book-to-market-, and momentum-matched control firms. Thus, large short positions are bearish signals.

Two necessary conditions for Miller's (1977) hypothesis are the existence of short sale constraints and dispersion of investor opinion. Boehme et al. (2006) examine Miller's hypothesis in a two-dimensional framework that accounts for differences of opinion and short sale constraints.¹⁰ The authors obtain daily short stock rebate and fee data from a major broker-dealer for the period from March 2001 to December 2002. Using a calendar-time methodology and the four-factor asset pricing model, the authors show portfolios with high short sale constraints are associated with lower abnormal returns. However, results are weaker when sorting by dispersion. Finally, when double sorting by constraint and dispersion, the portfolio with the highest dispersion and constraints has a 1-month abnormal return of -1.64% , while it is 0.49% for the portfolio with the highest dispersion but least constraint. These findings support the argument that firms subject to both high short sale constraints and high dispersion of investor opinion are likely to be overvalued.

Diether et al. (2009b) investigate short selling based on Reg SHO data. They measure daily stock-level shorting activity (*relss*) as the number of shorted shares divided by traded shares per day and show short sellers increase their shorting activity after periods of positive returns, on days with significant buying pressure, and on days with high levels of asymmetric information, which are consistent with three types of trading strategies: trading on short-term overreaction, acting as voluntary liquidity provider, and opportunistically providing risk-bearing services. Moreover, through portfolio analysis and regression framework, the authors show high short selling activity is associated with negative future abnormal returns.

Boehmer et al. (2010) take a new perspective by studying the performance of stocks with low short interest. The short interest data are obtained directly from the NYSE, Amex, and Nasdaq. Three heavily shorted and three lightly shorted portfolios are formed each month based on the short interest ratio (SIR) distribution in the previous month. The lightly shorted portfolios include stocks from the 1st, 5th, and 10th SIR percentiles, while the heavily shorted portfolios include stocks from the 99th, 95th, and 90th SIR percentiles. The authors are particularly interested in the three lightly shorted portfolios. Empirical results show the absence of short selling is a positive signal about future returns. For example, the SIR 1% portfolio realizes a 2% raw return or 1.4% adjusted return (alpha from the four-factor model of Fama and French (1993) and Carhart (1997)). Moreover, these positive future returns are larger in magnitude than the negative returns on heavily shorted stocks. Their findings cast doubt on existing theories about the impact of short sale constraints. Short sale constraints impede the incorporation of negative information in stock prices, thus result in overpricing (Miller 1977; Diamond and Verrecchia 1987). However, as there are no restrictions on going long, positive information should be incorporated in stock prices very quickly. The avoidance of short sellers for

¹⁰Previous empirical tests of Miller's hypothesis only focus on one of these two necessary conditions. For example, see Figlewski (1981) and Desai et al. (2002) for studies that examine the effect of short sale constraints on overvaluation; see Diether et al. (2002) and Danielsen and Sorescu (2001) for studies that examine the effect of dispersion of opinion on overvaluation.

undervalued stocks suggests that positive information is not incorporated in stock prices promptly, which undermines the critical asymmetry between the way good and bad news is revealed to the market. It seems that there are other frictions in the market, which lead to the predictability of future returns for stocks with low short interest level.

Engelberg et al. (2018) measure short selling risk by projecting the variance of loan fees (data from Markit) on several equity lending market characteristics and firm characteristics. Each month, equal-weighted quintile portfolios are formed based on the previous month's short selling risk. Stocks in the low (high) short selling risk quintile earn a monthly return of 0.58% (−0.49%) per month, suggesting short selling risk negatively predicts future returns. To control for the effect of short interest level, the authors perform 5×5 sorting, where stocks are first ranked by short interest level and then within each short interest quintiles, stocks are further sorted by short selling risk. The long-short portfolio returns range from 0.63% to 1.19%, so the effect of short selling risk is not subsumed by the well-established relation between short interest level and returns. Moreover, the authors document short selling risk is positively related to the Hou and Moskowitz (2005) measure of price delay, and negatively associated with short volume.

Bao et al. (2019) study whether managers have a tendency to disclose or withhold bad news through a proxy (residual short interest) for manager's private bad news. The novel approach tackles the challenge that researchers cannot observe the negative private information directly. The residual short interest contains negative information that is yet to be reflected in the stock price as it negatively predicts future abnormal stock returns. Moreover, the authors show that residual short interest is associated with insider trading and stock repurchases, indicating it indeed captures manager's negative private information. After validating residual short interest's effectiveness in proxying for negative private information, the authors further document a negative relation between bad-news disclosure and residual short interest, suggesting that managers withhold bad news in general.

In what follows, we review three papers on the effects of short selling on managerial decisions. Grullon et al. (2015) study the effects of short selling constraints on firm's investment and financing decisions. Short selling constraints can lead to overvaluation and overvaluation can cause overinvestment by artificially reducing the firm's cost of capital (Gilchrist et al. 2005). Thus, when short selling constraints are muted for pilot firms, we should observe lower investment. Consistent with this view, exploiting the regulatory change of Regulation SHO program, the authors find pilot firms with an increase in short selling activity exhibit larger reductions in investment and equity issuance. In the univariate difference-in-difference test, the estimator is −0.60 for capital expenditures, −0.52 for equity issuance, and −1.38 for debt issuance.

De Angelis et al. (2017) study short selling threat and the design of executive incentive contracts. Exploiting the natural experiment of Reg SHO program and the difference-in-difference approach, the authors show pilot firms whose short selling constraints are relaxed tend to increase the convexity of the compensation payoff. Further analysis shows that pilot firms do this by substituting away from restricted

stock and toward stock options in their new equity grants. Therefore, firms incentivize managers to alleviate the adverse effects of unrestrained short selling when facing the threat. The authors also find pilot firms adopt new antitakeover provisions.

Chang et al. (2019) examine the merger and acquisition (M&A) under the short selling threat. Managers face the temptation of pursuing private benefits when considering a deal, however, they are less likely to conduct value-destroying deals when there is a strong short selling threat. Examining 452 completed US domestic M&A from 2003 to 2012, the authors find acquirers with higher short interest gain higher M&A announcement returns, which lends support to the hypothesis. Further examination reveals that the effect is stronger when acquirers are more likely to be targets of subsequent hostile takeovers and when their managers' wealth is more linked to stock prices.

98.6 Future Research

98.6.1 CSR and Short Interest

As there is an increasing awareness of sustainability and corporate social responsibility (CSR), whether short sellers target firms with poor CSR performance is an open question and remains unexplored. Whether firms engage in CSR practices in the presence of increased threat from short selling is also an interesting avenue for future research. CSR is a broad concept that can take many forms depending on the types of company and industry. Using the data from China, Hou et al. (2019) document that a firm responds to the surge in short selling with an engagement in corporate philanthropic (CP) activities. Empirical evidence shows that a firm will use CP activities to divert public attention when it faces weaknesses such as the threat of short selling. However, the CP variable (whether or not a firm makes CP in a given year) they consider fails to capture the bigger dimension of CSR activities. Jain et al. (2016) find that short sellers avoid firms with high environmental, social, and governance (ESG) scores and tend to target firms with low ESG scores. Nevertheless, ESG also fails to capture the broad picture of CSR. An integration study between CSR and short selling is thus in great need.

98.6.2 The Information Role of Short Interest in Other Asset Markets

We notice that the information role of short interest in other asset markets is not fully understood. A smaller subset of papers focuses on this direction. For example, Kecskés et al. (2013) and Henry et al. (2015) find short interest can predict bond downgrades. Griffin et al. (2016) show short interest is a strong predictor of CDS returns. Meanwhile, a few studies focus on the substitution role of put option on short selling (Ackert and Athanassakos 2005; Grundy et al. 2012). We believe the role of short interest in the options market as well as other asset markets offers a promising avenue for further research.

98.6.3 Heterogeneity in Short Sellers

Short sellers are not homogeneous as the motivation for short selling varies in the first place. Brunnermeier and Oehmke (2014) study the behavior of predatory short sellers as their short selling pressure forces a financial institution to fire sell long-term assets to repay debt and satisfy the leverage constraints imposed by short-term creditors. Kelley and Tetlock (2017) find retail short selling predicts negative stock returns, and the return predictability is not subsumed by institutional short selling, which they attribute to the unique insights that retail short sellers have. Comerton-Forde et al. (2016b) find two distinct types of short sales by examining returns, order flows, and market conditions in the minutes before, during, and after NYSE and Nasdaq short sales, which they refer to as liquidity-supplying shorts and liquidity-demanding shorts. Liquidity-supplying short sellers trade when spreads are unusually wide, whereas liquidity-demanding short sellers trade when spreads are narrow and there is a decline in short-term price. Using the equity lending data of the Brazilian market, Chague et al. (2019) find that skilled short sellers present out-of-sample performance persistence while unskilled short sellers are likely to display the disposition effect.

Reed et al. (2019) differentiate the trading venues through which the short orders are executed, namely, exchange versus dark pools, which can answer the question regarding where short sellers exploit their information advantage. The authors find that the exchange short sales are more informative about future prices than dark pool short sales.

To summarize, different types of short sellers use different sets of information through different venues to exploit their information advantages, which warrant further investigation.

98.6.4 Endogeneity Problem

The endogeneity problem is unavoidable and present in many short selling studies. For example, Jain et al. (2019) find short interest decreases with an increase in accounting conservatism, while Jin et al. (2018) and Young (2016) document that firms will report more conservatively in the presence of short-selling threats.¹¹ In other words, conservatism determines the short selling in Jain et al. (2019), but is the consequence of short selling in Jin et al. (2018) and Young (2016). Short sellers target firms with poor financial reporting quality in the first place, and firms, in turn, respond with disciplined management and improved financial reporting quality, which will result in reverse causality.

Some studies employ the two-stage instrumental method to overcome or alleviate the endogeneity concern. However, the selection of an appropriate instrumental

¹¹ In a separate study, Mashruwala and Mashruwala (2018) find accounting conservatism alleviates the overvaluation for firms with the highest short selling constraints.

variable is a challenge. Hirshleifer et al. (2011) investigate whether short sellers target firms with high accruals, and the instrumental variable used in this study is institutional ownership, which acts as the proxy for the ease of short arbitrage. The use of institutional ownership as an instrument helps allay concern that the relation between accruals and short interest is driven by omitted variables rather than arbitrage activity. However, due to the monitoring function of institutional investors, the overall institutional ownership is not an appropriate instrument in most cases. For this reason, the literature (e.g., Massa et al. 2015; Meng et al. 2020) identifies an alternative instrumental variable, i.e., the ownership of exchange-traded funds (ETFs). ETFs have the lowest management fee among ETFs, brokerages, and banks, which are the main sources of lendable shares for the short selling market; short sellers thus have the highest incentive to borrow shares from ETFs. Moreover, ETFs are passive investors who typically track a stock market index and are less concerned about specific firm performance (Massa et al. 2015).

Meanwhile, many studies use difference-in-difference tests to tackle the endogeneity problem. For example, using the natural experiment of Reg SHO, Young (2016) find a decrease in short selling constraints causes a decrease in conditional conservatism in financial reporting. The regulatory change provides an ideal setting that frees the tests from the endogeneity concern. However, these studies are limited to the US, China, or Hong Kong markets because only these equity markets have experienced regulatory changes that affect the short selling activities.

98.6.5 International Evidence

Thus far, the cross-country studies are still quite limited. Existing studies mainly focus on a single country or market. While studies that use the US data account for the highest proportion of short sale research, we notice there are papers that investigate UK (Andrikopoulos et al. 2012; Au et al. 2009), China (Chang et al. 2014; Meng et al. 2017; Meng et al. 2020; Zhao et al. 2014; Hu et al. 2020; Ni and Zhu 2016; Chen et al. 2018), Canada (Ackert and Athanassakos 2005), Brazil (Chague et al. 2019), Australia (Comerton-Forde et al. 2016a; Aitken et al. 1998), Turkey (Sobaci et al. 2014), and Hong Kong market (Bai and Qin 2014, 2015; Chang et al. 2007; Chen and Rhee 2010; Choy and Zhang 2019; McKenzie and Henry 2012). A few papers conduct analysis globally; for example, Beber and Pagano (2013) find short sale bans or regulatory constraints on short selling are harmful to the liquidity and will slow the price discovery process in 30 countries in their sample. Bris et al. (2007) study the data from 46 countries and document that allowing short sales will improve market efficiency. Saffi and Sigurdsson (2011) study a global dataset from 26 countries and find stocks with higher short sale constraints have lower price efficiency; most importantly, relaxing short sale constraints is not associated with an increase in market instability. Massa et al. (2015) explore the relationship between short selling and earnings management using a global sample from 33 countries. Jiang et al. (2020b) extend Massa et al.'s (2015) research by showing that a short-selling threat constrains real earnings management

internationally. The development of financial markets at different levels across countries, different roles of institutions play in those markets, as well as heterogeneities of investors' financial literacy warrant further investigation.

98.7 Conclusion

Short sales have profound impacts on asset price behaviors, market equilibrium, and corporate finance and investment decisions. As such, short sellers' activity has received a considerable amount of attention from academics and policy makers. In this chapter, we review the current state of research on short sales and suggest several future directions for this line of research in order to improve our understanding of the role of short sellers in capital market development and in influencing corporate decisions.

A clear consensus in the literature is the finding that short sellers are knowledgeable traders who are capable of collecting both public and private information to facilitate their trading decisions. Short selling activity has significant impacts on information dissemination, price discovery, market efficiency, and corporate decision-making. Restricting short sales has adverse effects on pricing efficiency and price discovery.

We review the literature on the information content in short interest, information advantages of short sellers, and firms' reactions to the short selling threat. In addition, we discuss the data and metrics commonly used in short sales studies, and summarize important empirical findings. We highlight some limitations of current research and areas that require further investigation to improve our understanding of the role of short sellers. In particular, we need to understand the sources of information possessed by short sellers and the effects of short selling in different environments (e.g., investment sentiment and financial distress).

Although several theoretical models have been developed to explain the effects of short selling, our knowledge for the role of short sellers in affecting corporate governance and decisions remains quite limited, and many issues related to short sales are unresolved. It is therefore important for financial researchers to develop better models to explain these unresolved issues and provide guided empirical analysis to further improve our understanding of the role of short selling in financial markets and corporate managers' decision-making.

References

- Ackert, Lucy F., and George Athanassakos. 2005. The relationship between short interest and stock returns in the Canadian market. *Journal of Banking & Finance* 29 (7): 1729–1749.
- Aitken, Michael J., Alex Frino, Michael S. McCorry, and Peter L. Swan. 1998. Short sales are almost instantaneously bad news: Evidence from the Australian Stock Exchange. *Journal of Finance* 53 (6): 2205–2223.
- Akbas, Ferhat, Ekkehart Boehmer, Bilal Erturk, and Sorin Sorescu. 2017. Short interest, returns, and unfavorable fundamental information. *Financial Management* 46 (2): 455–486.

- Allen, Franklin, and Douglas Gale. 1992. Stock-price manipulation. *Review of Financial Studies* 5 (3): 503–529.
- Andrikopoulos, Panagiotis, James Clunie, and Antonios Siganos. 2012. UK short selling activity and firm performance. *Journal of Business Finance & Accounting* 39 (9–10): 1403–1417.
- Angelis, De, Gustavo Grullon David, and Sébastien Michenaud. 2017. The effects of short-selling threats on incentive contracts: Evidence from an experiment. *Review of Financial Studies* 30 (5): 1627–1659.
- Asquith, Paul, Parag A. Pathak, and Jay R. Ritter. 2005. Short interest, institutional ownership, and stock returns. *Journal of Financial Economics* 78 (2): 243–276.
- Asquith, Paul, Rebecca Oman, and Christopher Safaya. 2010. Short sales and trade classification algorithms. *Journal of Financial Markets* 13 (1): 157–173.
- Atmaz, Adem, and Suleyman Basak. 2019. Option prices and costly short-selling. *Journal of Financial Economics* 134 (1): 1–28.
- Au, Andrea S., John A. Doukas, and Zhan Onayev. 2009. Daily short interest, idiosyncratic risk, and stock returns. *Journal of Financial Markets* 12 (2): 290–316.
- Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov. 2007. Momentum and credit rating. *Journal of Finance* 62 (5): 2503–2520.
- Bai, Min, and Yafeng Qin. 2014. Short-sales constraints and liquidity change: Cross-sectional evidence from the Hong Kong market. *Pacific-Basin Finance Journal* 26: 98–122.
- . 2015. Short sales constraints and price adjustments to earnings announcements: Evidence from the Hong Kong market. *International Review of Financial Analysis* 42: 304–315.
- Baker, Malcolm, and Serkan Savaşoglu. 2002. Limited arbitrage in mergers and acquisitions. *Journal of Financial Economics* 64 (1): 91–115.
- Bao, Dichu, Yongtae Kim, G. Mujtaba Mian, and Lixin Su. 2019. Do managers disclose or withhold bad news? Evidence from short interest. *The Accounting Review* 94 (3): 1–26.
- Beber, Alessandro, and Marco Pagano. 2013. Short-selling bans around the world: Evidence from the 2007–09 crisis. *Journal of Finance* 68 (1): 343–381.
- Beneish, Messod D., Charles M.C. Lee, and D. Craig Nichols. 2015. In short supply: Short-sellers and stock returns. *Journal of Accounting and Economics* 60 (2–3): 33–57.
- Berkman, Henk, and Michael D. McKenzie. 2012. Earnings announcements: Good news for institutional investors and short sellers. *Financial Review* 47 (1): 91–113.
- Berkman, Henk, Michael D. McKenzie, and Patrick Verwijmeren. 2017. Hole in the wall: Informed short selling ahead of private placements. *Review of Finance* 21 (3): 1047–1091.
- Blau, Benjamin M., Tyler J. Brough, Jason L. Smith, and Nathaniel M. Stephens. 2013. An examination of short-selling activity surrounding auditor changes. *Journal of Accounting, Auditing and Finance* 28 (4): 348–368.
- Blocher, Jesse, Adam V. Reed, and Edward D. Van Wesep. 2013. Connecting two markets: An equilibrium framework for shorts, longs, and stock loans. *Journal of Financial Economics* 108 (2): 302–322.
- Boehme, Rodney D., Bartley R. Danielsen, and Sorin M. Sorescu. 2006. Short-sale constraints, differences of opinion, and overvaluation. *Journal of Financial and Quantitative Analysis* 41: 455–487.
- Boehmer, Ekkehart, and Juan Wu. 2013. Short selling and the price discovery process. *Review of Financial Studies* 26 (2): 287–322.
- Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang. 2008. Which shorts are informed? *Journal of Finance* 63 (2): 491–527.
- Boehmer, Ekkehart, Zsuzsa R. Huszar, and Bradford D. Jordan. 2010. The good news in short interest. *Journal of Financial Economics* 96 (1): 80–97.
- Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang. 2013. Shackling short sellers: The 2008 shorting ban. *Review of Financial Studies* 26 (6): 1363–1400.
- Boulton, Thomas J., Scott B. Smart, and Chad J. Zutter. 2020. Worldwide short selling regulations and IPO underpricing. *Journal of Corporate Finance* 62: 101596.
- Brent, Averil, Dale Morse, and E. Kay Stice. 1990. Short interest: Explanations and tests. *Journal of Financial and Quantitative Analysis* 25: 273–289.

- Bris, Arturo, William N. Goetzmann, and Ning Zhu. 2007. Efficiency and the bear: Short sales and markets around the world. *Journal of Finance* 62 (3): 1029–1079.
- Brunnermeier, Markus K., and Martin Oehmke. 2014. Predatory short selling. *Review of Finance* 18 (6): 2153–2195.
- Carhart, Mark M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52 (1): 57–82.
- Chague, Fernando, Rodrigo De-Losso, Alan De Genaro, and Bruno Giovannetti. 2017. Well-connected short-sellers pay lower loan fees: A market-wide analysis. *Journal of Financial Economics* 123 (3): 646–670.
- Chague, Fernando, Rodrigo De-Losso, and Bruno Giovannetti. 2019. The short-selling skill of institutions and individuals. *Journal of Banking & Finance* 101: 77–91.
- Chakrabarty, Bidisha, and Andriy Shkilko. 2013. Information transfers and learning in financial markets: Evidence from short selling around insider sales. *Journal of Banking & Finance* 37 (5): 1560–1572.
- Chang, Eric C., Joseph W. Cheng, and Yu. Yinghui. 2007. Short-sales constraints and price discovery: Evidence from the Hong Kong market. *Journal of Finance* 62 (5): 2097–2121.
- Chang, Eric C., Yan Luo, and Jinjuan Ren. 2014. Short-selling, margin-trading, and price efficiency: Evidence from the Chinese market. *Journal of Banking & Finance* 48: 411–424.
- Chang, Eric C., Tse-Chun Lin, and Xiaorong Ma. 2019. Does short-selling threat discipline managers in mergers and acquisitions decisions? *Journal of Accounting and Economics* 68 (1): 101223.
- Chen, Crystal Xiaobei, and S. Ghon Rhee. 2010. Short sales and speed of price adjustment: Evidence from the Hong Kong stock market. *Journal of Banking & Finance* 34 (2): 471–483.
- Chen, Shenglan, Bingxuan Lin, Rui Lu, and Hui Ma. 2018. The disciplinary effects of short sales on controlling shareholders. *Journal of Empirical Finance* 46: 56–76.
- Chen, Hang, Yushu Zhu, and Chang Liang. 2019. Short-selling constraints and corporate payout policy. *Accounting and Finance* 59 (4): 2273–2305.
- Chen, Xia, Qiang Cheng, Ting Luo, and Heng Yue. 2020. Short sellers and long-run management forecasts. *Contemporary Accounting Research* 37 (2): 802–828.
- Cheng, Lee-Young, Zhipeng Yan, Yan Zhao, and Wei-Fang Chang. 2012. Short selling activity, price efficiency and fundamental value of IPO stocks. *Pacific-Basin Finance Journal* 20 (5): 809–824.
- Choy, Siu Kai, and Hua Zhang. 2019. Public news announcements, short-sale restriction and informational efficiency. *Review of Quantitative Finance and Accounting* 52 (1): 197–229.
- Christensen, Theodore E., Michael S. Drake, and Jacob R. Thornock. 2014. Optimistic reporting and pessimistic investing: Do pro forma earnings disclosures attract short sellers? *Contemporary Accounting Research* 31 (1): 67–102.
- Christophe, Stephen E., Michael G. Ferri, and James J. Angel. 2004. Short-selling prior to earnings announcements. *Journal of Finance* 59 (4): 1845–1876.
- Christophe, Stephen E., Michael G. Ferri, and Jim Hsieh. 2010. Informed trading before analyst downgrades: Evidence from short sellers. *Journal of Financial Economics* 95 (1): 85–106.
- Comerton-Forde, Carole, Binh Huu Do, Philip Gray, and Tom Manton. 2016a. Assessing the information content of short-selling metrics using daily disclosures. *Journal of Banking & Finance* 64: 188–204.
- Comerton-Forde, Carole, Charles M. Jones, and Tălis J. Putniņš. 2016b. Shorting at close range: A tale of two types. *Journal of Financial Economics* 121 (3): 546–568.
- Curtis, Asher, and Neil L. Fargher. 2014. Does short selling amplify price declines or align stocks with their fundamental values? *Management Science* 60 (9): 2324–2340.
- D’avolio, Gene. 2002. The market for borrowing stock. *Journal of Financial Economics* 66 (2–3): 271–306.
- Danielsen, Bartley R., and Sorin M. Sorescu. 2001. Why do option introductions depress stock prices? A study of diminishing short sale constraints. *Journal of Financial and Quantitative Analysis* 36: 451–484.

- Dechow, Patricia M., Richard G. Sloan, and Amy P. Sweeney. 1996. Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research* 13 (1): 1–36.
- Dechow, Patricia M., Amy P. Hutton, Lisa Meulbroeck, and Richard G. Sloan. 2001. Short-sellers, fundamental analysis, and stock returns. *Journal of Financial Economics* 61 (1): 77–106.
- Deng, Xiaohu, Lei Gao, and Jeong-Bon Kim. 2020. Short-sale constraints and stock price crash risk: Causal evidence from a natural experiment. *Journal of Corporate Finance* 60: 101498.
- Desai, Hemang, Kevin Ramesh, S. Ramu Thiagarajan, and Bala V. Balachandran. 2002. An investigation of the informational role of short interest in the Nasdaq market. *Journal of Finance* 57 (5): 2263–2287.
- Desai, Hemang, Srinivasan Krishnamurthy, and Kumar Venkataraman. 2006. Do short sellers target firms with poor earnings quality? Evidence from earnings restatements. *Review of Accounting Studies* 11 (1): 71–90.
- Deshmukh, Sanjay, Keith Jacks Gamble, and Keith M. Howe. 2015. Short selling and firm operating performance. *Financial Management* 44 (1): 217–236.
- Diamond, Douglas W., and Robert E. Verrecchia. 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics* 18 (2): 277–311.
- Diether, Karl B., Christopher J. Malloy, and Anna Scherbina. 2002. Differences of opinion and the cross section of stock returns. *Journal of Finance* 57 (5): 2113–2141.
- Diether, Karl B., Kuan-Hui Lee, and Ingrid M. Werner. 2009a. It's SHO time! Short-sale price tests and market quality. *Journal of Finance* 64 (1): 37–73.
- . 2009b. Short-sale strategies and return predictability. *Review of Financial Studies* 22 (2): 575–607.
- Drake, Michael S., Lynn Rees, and Edward P. Swanson. 2011. Should investors follow the prophets or the bears? Evidence on the use of public information by analysts and short sellers. *The Accounting Review* 86 (1): 101–130.
- Drake, Michael S., Linda A. Myers, Susan Scholz, and Nathan Y. Sharp. 2015. Short selling around restatement announcements: When do bears pounce? *Journal of Accounting, Auditing and Finance* 30 (2): 218–245.
- Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen. 2002. Securities lending, shorting, and pricing. *Journal of Financial Economics* 66 (2–3): 307–339.
- Dyl, Edward A. 1978. Short selling and the capital gains tax. *Financial Analysts Journal* 34 (2): 61–64.
- Edwards, Amy K., and Kathleen Weiss Hanley. 2010. Short selling in initial public offerings. *Journal of Financial Economics* 98 (1): 21–39.
- Edwards, Alexander, Casey Schwab, and Terry Shevlin. 2016. Financial constraints and cash tax savings. *The Accounting Review* 91 (3): 859–881.
- Engelberg, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg. 2012. How are shorts informed?: Short sellers, news, and information processing. *Journal of Financial Economics* 105 (2): 260–278.
- . 2018. Short-selling risk. *Journal of Finance* 73 (2): 755–786.
- Evans, Richard B., Christopher C. Geczy, David K. Musto, and Adam V. Reed. 2009. Failure is an option: Impediments to short selling and options prices. *Review of Financial Studies* 22 (5): 1955–1980.
- Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.
- Fang, Vivian W., Allen H. Huang, and Jonathan M. Karpoff. 2016. Short selling and earnings management: A controlled experiment. *Journal of Finance* 71 (3): 1251–1294.
- Figlewski, Stephen. 1981. The informational effects of restrictions on short sales: Some empirical evidence. *Journal of Financial and Quantitative Analysis* 16: 463–476.
- Frino, Alex, Steven Lecce, and Andrew Lepone. 2011. Short-sales constraints and market quality: Evidence from the 2008 short-sales bans. *International Review of Financial Analysis* 20 (4): 225–236.

- Geczy, Christopher C., David K. Musto, and Adam V. Reed. 2002. Stocks are special too: An analysis of the equity lending market. *Journal of Financial Economics* 66 (2–3): 241–269.
- Gilchrist, Simon, Charles P. Himmelberg, and Gur Huberman. 2005. Do stock price bubbles influence corporate investment? *Journal of Monetary Economics* 52 (4): 805–827.
- Goldstein, Itay, and Alexander Guembel. 2008. Manipulation and the allocational role of prices. *Review of Economic Studies* 75 (1): 133–164.
- Gong, Rong. 2020. Short selling threat and corporate financing decisions. *Journal of Banking & Finance* 118: 105853.
- Griffin, Paul A., Hyun A. Hong, and Jeong-Bon Kim. 2016. Price discovery in the CDS market: The informational role of equity short interest. *Review of Accounting Studies* 21 (4): 1116–1148.
- Grullon, Gustavo, Sébastien Michenaud, and James P. Weston. 2015. The real effects of short-selling constraints. *Review of Financial Studies* 28 (6): 1737–1767.
- Grundy, Bruce D., Bryan Lim, and Patrick Verwijmeren. 2012. Do option markets undo restrictions on short sales? Evidence from the 2008 short-sale ban. *Journal of Financial Economics* 106 (2): 331–348.
- Guo, Xu, and Chunchi Wu. 2019. Short interest, stock returns and credit ratings. *Journal of Banking & Finance* 108: 105617.
- Guo, Savannah Yuanyuan, Sabrina Chi, and Kirsten A. Cook. 2018. Short selling and corporate tax avoidance. In *Advances in taxation*. Emerald Publishing Limited. 25: 1–28.
- Hasan, Iftexhar, Nadia Massoud, Anthony Saunders, and Keke Song. 2015. Which financial stocks did short sellers target in the subprime crisis? *Journal of Banking & Finance* 54: 87–103.
- Henry, Tyler R., Darren J. Kisgen, and Juan Julie Wu. 2015. Equity short selling and bond rating downgrades. *Journal of Financial Intermediation* 24 (1): 89–111.
- Hirshleifer, David, Siew Hong Teoh, and Jeff Jiewei Yu. 2011. Short arbitrage, return asymmetry, and the accrual anomaly. *Review of Financial Studies* 24 (7): 2429–2461.
- Ho, Kim Wai. 1996. Short-sales restrictions and volatility: The case of the Stock Exchange of Singapore. *Pacific-Basin Finance Journal* 4 (4): 377–391.
- Hobbs, Jeffrey, Terrill R. Keasler, and Chris R. McNeil. 2012. Short selling behavior and mad money. *Financial Review* 47 (1): 65–89.
- Hong, Harrison, and Jeremy C. Stein. 2003. Differences of opinion, short-sales constraints, and market crashes. *Review of Financial Studies* 16 (2): 487–525.
- Hong, Harrison, Jose Scheinkman, and Wei Xiong. 2006. Asset float and speculative bubbles. *Journal of Finance* 61 (3): 1073–1117.
- Hou, Kewei, and Tobias J. Moskowitz. 2005. Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies* 18 (3): 981–1020.
- Hou, Deshuai, Qingbin Meng, Kai Zhang, and Kam C. Chan. 2019. Motives for corporate philanthropy propensity: Does short selling matter? *International Review of Economics and Finance* 63: 24–36.
- Hu, Ning, Siqi Lu, Tao Ma, and Jianfang Ye. 2020. Short-selling and cost of equity: Evidence from China. *Accounting and Finance* 60 (4): 3681–3707.
- Huszár, Zsuzsa R., Ruth S.K. Tan, and Weina Zhang. 2017. Do short sellers exploit industry information? *Journal of Empirical Finance* 41: 118–139.
- Jain, Archana, Pankaj K. Jain, and Zabiollah Rezaee. 2016. Value-relevance of corporate social responsibility: Evidence from short selling. *Journal of Management Accounting Research* 28 (2): 29–52.
- Jain, Archana, Chinmay Jain, and Ashok Robin. 2019. Does accounting conservatism deter short sellers? *Review of Quantitative Finance and Accounting* 54: 1–26.
- Jarrow, Robert. 1980. Heterogeneous expectations, restrictions on short sales, and equilibrium asset prices. *Journal of Finance* 35 (5): 1105–1113.
- Jiang, Danling, David R. Peterson, and James S. Doran. 2014. Short-sale constraints and the idiosyncratic volatility puzzle: An event study approach. *Journal of Empirical Finance* 28: 36–59.
- Jiang, Haiyan, Ahsan Habib, and Mostafa Monzur Hasan. 2020a. \Short selling: A review of the literature and implications for future research. *The European Accounting Review* 1–31.

- Jiang, Haiyan, Yafeng Qin, and Min Bai. 2020b. Short-selling threats and real earnings management – International evidence. *Journal of International Accounting Research* 19 (2): 117–140.
- Jin, Zhi, Bingxuan Lin, Yang Xue, and Ting Zhang. 2018. Accounting conservatism and short selling: Evidence from China. *Journal of Business Finance & Accounting* 45 (3–4): 352–394.
- Jones, Charles M. 2012. Shorting restrictions: Revisiting the 1930s. *Financial Review* 47 (1): 1–35.
- Jones, Charles M., and Owen A. Lamont. 2002. Short-sale constraints and stock returns. *Journal of Financial Economics* 66 (2–3): 207–239.
- Karpoff, Jonathan M., and Xiaoxia Lou. 2010. Short sellers and financial misconduct. *Journal of Finance* 65 (5): 1879–1913.
- Keckskés, Ambrus, Sattar A. Mansi, and Andrew Jianzhong Zhang. 2013. Are short sellers informed? Evidence from the bond market. *The Accounting Review* 88 (2): 611–639.
- Kelley, Eric K., and Paul C. Tetlock. 2017. Retail short selling and stock prices. *Review of Financial Studies* 30 (3): 801–834.
- Khan, Mozaffar, and Hai Lu. 2013. Do short sellers front-run insider sales? *The Accounting Review* 88 (5): 1743–1768.
- Lamont, Owen A. 2012. Go down fighting: Short sellers vs. firms. *Review of Asset Pricing Studies* 2 (1): 1–30.
- Lamont, Owen A., and Jeremy C. Stein. 2004. Aggregate short interest and market valuations. *American Economic Review* 94 (2): 29–32.
- Lee, Eunju, and Natalia Piqueira. 2017. Short selling around the 52-week and historical highs. *Journal of Financial Markets* 33: 75–101.
- Li, Yinghua, and Liandong Zhang. 2015. Short selling pressure, stock price behavior, and management forecast precision: Evidence from a natural experiment. *Journal of Accounting Research* 53 (1): 79–117.
- Lin, Tse-Chun, and Xiaolong Lu. 2016. How do short-sale costs affect put options trading? Evidence from separating hedging and speculative shorting demands. *Review of Finance* 20 (5): 1911–1943.
- Liu, Ming, Tongshu Ma, and Yan Zhang. 2012. Are short sellers informed? Evidence from the 2007–2008 subprime mortgage crisis. *Financial Review* 47 (1): 199–218.
- Ljungqvist, Alexander, and Wenlan Qian. 2016. How constraining are limits to arbitrage? *Review of Financial Studies* 29 (8): 1975–2028.
- Lynch, Andrew, Biljana Nikolic, Xuemin Sterling Yan, and Yu. Han. 2014. Aggregate short selling, commonality, and stock market returns. *Journal of Financial Markets* 17: 199–229.
- Mashruwala, Christina, and Shamin Mashruwala. 2018. Does accounting conservatism reduce overpricing caused by short-sales constraints? *Contemporary Accounting Research* 35 (4): 2160–2190.
- Massa, Massimo, Bohui Zhang, and Hong Zhang. 2015. The invisible hand of short selling: Does short selling discipline earnings management? *Review of Financial Studies* 28 (6): 1701–1736.
- McKenzie, Michael, and Ólan T. Henry. 2012. The determinants of short selling: Evidence from the Hong Kong equity market. *Accounting and Finance* 52: 183–216.
- Meng, Qingbin, Ying Li, Xuanyu Jiang, and Kam C. Chan. 2017. Informed or speculative trading? Evidence from short selling before star and non-star analysts' downgrade announcements in an emerging market. *Journal of Empirical Finance* 42: 240–255.
- Meng, Qingbin, Xinyu Li, Kam C. Chan, and Shenghao Gao. 2020. Does short selling affect a firm's financial constraints? *Journal of Corporate Finance* 60: 101531.
- Miller, Edward M. 1977. Risk, uncertainty, and divergence of opinion. *Journal of Finance* 32 (4): 1151–1168.
- Nagel, Stefan. 2005. Short sales, institutional investors and the cross-section of stock returns. *Journal of Financial Economics* 78 (2): 277–309.
- Ni, Xiaoran, and Weikang Zhu. 2016. Short-sales and stock price crash risk: Evidence from an emerging market. *Economics Letters* 144: 22–24.
- Park, KoEun. 2017. Earnings quality and short selling: Evidence from real earnings management in the United States. *Journal of Business Finance & Accounting* 44 (9–10): 1214–1240.

- Pownall, Grace, and Paul J. Simko. 2005. The information intermediary role of short sellers. *The Accounting Review* 80 (3): 941–966.
- Purnanandam, Amiyatosh, and H. Nejat Seyhun. 2018. Do short sellers trade on private information or false information? *Journal of Financial and Quantitative Analysis* 53 (3): 997–1023.
- Rapach, David E., Matthew C. Ringgenberg, and Guofu Zhou. 2016. Short interest and aggregate stock returns. *Journal of Financial Economics* 121 (1): 46–65.
- Reed, Adam V. 2013. Short selling. *Annual Review of Financial Economics* 5 (1): 245–258.
- . 2015. Connecting supply, short-sellers and stock returns: Research challenges. *Journal of Accounting and Economics* 60 (2–3): 97–103.
- Reed, Adam V., Mehrdad Samadi, and Jonathan S. Sokobin. 2019. Shorting in broad daylight: Short sales and venue choice. *Journal of Financial and Quantitative Analysis*: 18–27.
- Saffi, Pedro A.C., and Kari Sigurdsson. 2011. Price efficiency and short selling. *Review of Financial Studies* 24 (3): 821–852.
- Senchack, A.J., Jr., and Laura T. Starks. 1993. Short-sale restrictions and market reaction to short-interest announcements. *Journal of Financial and Quantitative Analysis* 28: 177–194.
- Seneca, Joseph J. 1967. Short interest: bearish or bullish? *Journal of Finance* 22 (1): 67–70.
- Shi, Jian, Junbo Wang, and Ting Zhang. 2017. Are short sellers informed? Evidence from credit rating agency announcements. *Journal of Financial Research* 40 (2): 179–221.
- Shkilko, Andriy, Bonnie Van Ness, and Robert Van Ness. 2012. Short selling and intraday price pressures. *Financial Management* 41 (2): 345–370.
- Shyu, Yih-Wen, Kam C. Chan, and Hsin-Yu Liang. 2018. Spillovers of price efficiency and informed trading from short sales to margin purchases in absence of uptick rule. *Pacific-Basin Finance Journal* 50: 163–183.
- Sloan, Richard G. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71: 289–315.
- Sobaci, Cihat, Ahmet Sensoy, and Mutahhar Erturk. 2014. Impact of short selling activity on market dynamics: Evidence from an emerging market. *Journal of Financial Stability* 15: 53–62.
- Stambaugh, Robert F., Jianfeng Yu, and Yuan Yu. 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance* 70 (5): 1903–1948.
- Takahashi, Hidetomo. 2010. Short-sale inflow and stock returns: Evidence from Japan. *Journal of Banking & Finance* 34 (10): 2403–2412.
- Thornock, Jacob. 2013. The effects of dividend taxation on short selling and market quality. *The Accounting Review* 88 (5): 1833–1856.
- Young, Alex. 2016. Capital market frictions and conservative reporting: Evidence from short selling constraints. *Finance Research Letters* 17: 227–234.
- Zhao, Zhongkuang, Shuqi Li, and Heping Xiong. 2014. Short sale constraints, disperse pessimistic beliefs and market efficiency – Evidence from the Chinese stock market. *Economic Modelling* 42: 333–342.



Simultaneous Equation Models for Financial Planning and Forecasting

99

Cheng-Few Lee, Alice C. Lee, and Wen-chi Yeh

Abstract

This chapter first discusses Warren and Shelton (J Finance, 26:1123–1142, 2007) and Francis and Rowell's (Financ Manag 7:29–44, 1978) simultaneous equation models for financial planning and forecasting in detail. Then we use Johnson & Johnson's 2018 financial data in terms of Excel programming to demonstrate how Warren and Shelton's model can be used for financial analysis, planning, and forecasting. Data from 2019 and 2020 is used to evaluate the performance of Warren and Shelton's model. In addition, sensitivity analysis is used to demonstrate how the change of growth rate, debt equity ratio, and P/E ratio can affect the price per share for Johnson & Johnson.

Keywords

Simultaneous equation · Financial planning and forecasting · Sustainable growth rate · Debt equity ratio · Payout ratio · P/E ratio · Sensitivity analysis

C.-F. Lee (✉)

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: cflee@business.rutgers.edu

A. C. Lee

Center for PBBEF Research, Phoenix, AZ, USA

W.-c. Yeh

Rutgers University, New Brunswick, NJ, USA

e-mail: [wenchi.yeh@business.rutgers.edu](mailto:wENCHI.YEH@BUSINESS.RUTGERS.EDU)

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_102

2341

99.1 Introduction

This chapter covers alternative financial planning models and their use in financial analysis and decision-making. The approach taken in this chapter gives the student an opportunity to combine information (accounting, market, and economic), theory, and methodology (regression and Excel programming).

We begin by presenting the procedure for financial planning and analysis in Sect. 99.2. This is followed by a discussion of the Warren and Shelton algebraic simultaneous equations planning model in Sect. 99.3. Section 99.4 discusses the importance of sensitivity analysis and its application to Warren and Shelton's financial planning model, Sect. 99.5 discusses the Francis and Rowell model, and Sect. 99.6 offers a summary of the chapter. Appendix A is a description of parameter inputs used to forecast Johnson & Johnson's financial statements and share price. Appendix B shows the procedure of how to use Excel to implement the FinPlan program.

99.2 Procedures for Financial Planning and Analysis

Before discussing the various financial planning models, we must first be sure of our understanding of what the financial planning process is all about. Otherwise, we run the risk of too narrowly defining financial planning as simply data gathering and running computer programs. In reality, financial planning involves a *process* of analyzing alternative dividend, financing, and investment strategies, forecasting their outcome and impact within various economic environments, and then deciding how much risk to take on and which projects to pursue. Thus, financial planning models are merely tools to improve forecasting as well as to help managers better understand the interactions of dividend, financing, and investment decisions.

More formally, we can outline the financial planning process as follows:

1. *Utilize* the existing set of economic, legal, accounting, marketing, and company policy information.
2. *Analyze* the interactions of the dividend, financing, and investment choices open to the firm.
3. *Forecast* the future consequences of present decisions to avoid unexpected events as well as to aid in understanding the interaction of present and future decisions.
4. *Decide* which alternatives the firm should undertake, the explicit outline for which is contained in the financial plan.
5. *Evaluate* the subsequent outcome of these decisions once they are implemented against the objectives set forth in the financial plan.

So where does the financial planning model come in? To clarify its role in this process, look at Fig. 1, which presents a flowchart of a financial planning model. The

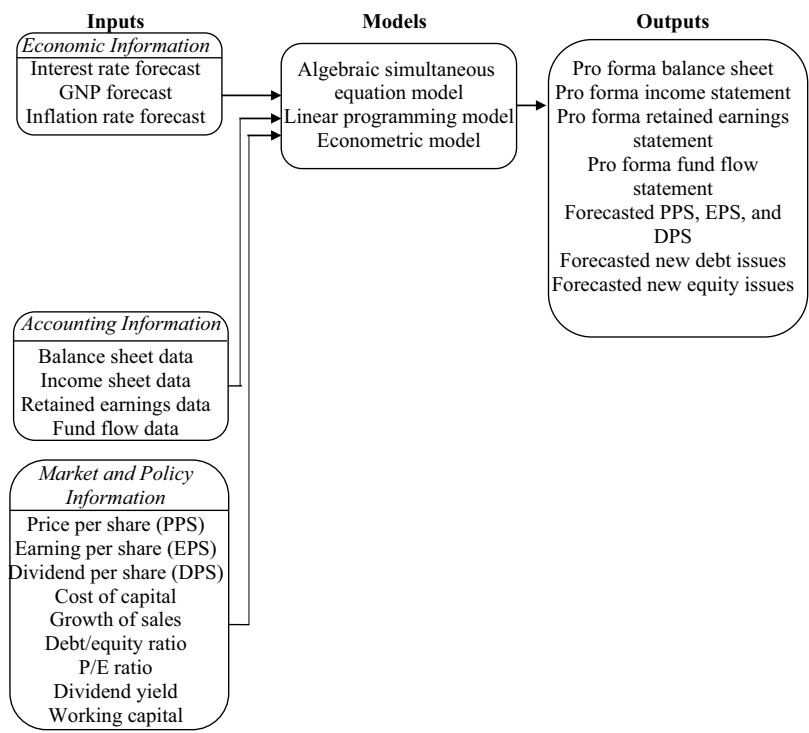


Fig. 1 Inputs, models, and outputs for financial planning and forecasting Models

inputs to the model are economic and accounting information and market and policy information. Two alternative financial planning, analysis, and forecasting models are (1) Warren and Shelton’s model and (2) Francis and Rowell’s model.¹ The outputs of the financial planning and forecasting model are pro forma financial statements, forecasted PPS, EPS, and DPS, new equity issued, and new debt issued. Essentially, the benefit of the model is to efficiently and effectively handle the analysis of information and its interactions with the forecasting of future consequences within the planning process.

Hence, the financial planning model efficiently improves the depth and breadth of the information the financial manager uses in the decision-making process. Moreover, before the finalized plan is implemented, an evaluation of how well subsequent performance stands up to the financial plan provides additional input for future planning actions.

A key to the value of any financial planning model is how it is formulated and constructed. That is, the credibility of the model’s output depends on the underlying assumptions and particular financial theory the model is based on, as well as its ease

¹This chapter discusses two alternative financial planning models. The simultaneous equation model can be found in Lee and Lee’s (2017), Chap. 24.

of use for the financial planner. Because of its potentially great impact on the financial planning process and, consequently, on the firm's future, the particular financial planning model to be used must be chosen carefully. Specifically, we can state that a useful financial planning model should have the following characteristics:

1. The model results and assumptions should be credible.
2. The model should be flexible so that it can be adapted and expanded to meet a variety of circumstances.
3. The model should improve on current practice in a technical or performance sense.
4. The model inputs and outputs should be comprehensible to the user without extensive additional knowledge or training.
5. The model should take into account the interrelated investment, financing, dividend, and production decisions and their effect on the firm's market value.
6. The model should be fairly simple for the user to operate without extensive intervention of nonfinancial personnel and tedious formulation of the input.

On the basis of these guidelines, we now present and discuss the simultaneous equations, linear programming, and econometric financial planning models, which can be used for financial planning and analysis.

99.3 Percentage of Sales Method for Financial Planning and Forecasting

Ross et al. (2019) and Lee et al. (1997) have shown how percentage sales method can be used to develop financial analysis, planning, and forecasting. Financial planning is the process of analyzing alternative investment, financing, and dividend strategies in the context of various potential economic environments. Planning involves forecasting both the outcomes of different strategies and their risks. Thus, financial planning models are tools to help managers improve their forecasts of important accounts of financial statements and better understand the interactions of investment, financing, and dividend decisions.

In developing a long-term financial plan, these three decisions (policies) can be described more explicitly as follows.

1. *The firm's investment decision.* This refers to the amount of cash needed for the firm's investment in a new asset (it is also called the *capital budgeting decision*). In addition, it also refers to the amount of working capital needed on an ongoing basis (also referred to as the *working capital decision*).
2. *The firm's financing decision.* This refers to new borrowing or new equity issued for financing the firm's investment in new assets. This decision is influenced by the degree of financial leverage the firm chooses to employ and how it plans to raise the necessary new funds.

3. *The firm's dividend decision.* This refers to the amount of cash the firm thinks is necessary and appropriate to pay equity holders as cash dividends.

At the most basic level, a planning model is a tool that uses inputs supplied by managers in the form of economic, accounting, market, and policy information. Depending on its sophistication, the output of the financial planning and forecasting model may include simple estimates of growth rates or complex pro forma financial statements, with projected financial statement data, and forecasts of stock price, earnings per share, dividends per share, new equity issues, and new debt issues.

The procedure of using percentage sales methods for financial planning and forecasting can be briefly described as follows:

1. Calculate by dividing company sales into both balance sheet and income statement to obtain standardized financial statements.
2. Forecast growth rate and outside financing.
3. Forecast external financing needs.
4. Combine information from steps 1, 2, 3 to obtain pro forma financial statements.
5. Use pro forma financial statements to perform future financial planning and forecasting.

99.4 The Algebraic Simultaneous Equations Approach to Financial Planning and Analysis

In this section, we present the financial planning approach of Warren and Shelton (1971), which is based on a *simultaneous equations* concept. The model, called FINPLAN, deals with overall corporate financial planning as opposed to just single planning, such as capital budgeting. The objective of the FINPLAN model is not to optimize anything, but, rather, to serve as a tool to provide relevant information to the decision-maker. One of the strengths of this planning model, in addition to its construction, is that it allows the user to simulate the financial impacts of changing assumptions regarding such variables as sales, operating ratios, price-to-earnings ratios, retention rates, and debt-to-equity ratios.

The advantage of utilizing a simultaneous equation structure to represent a firm's investment, financing, production, and dividend policies is the enhanced ability for the interaction of these decision-making areas. The Warren and Shelton (WS) model is a system of 20 equations, which are listed in Table 1. These equations are segmented into distinct subgroups corresponding to sales, investment, financing, and per share (return to investors) data. The flowchart describing the interrelationships of the equations is shown in Fig. 2.

The key concepts of the interaction of investment, financing, and dividends are the basis of the FINPLAN model, which we now consider in some detail. First, we discuss the inputs to the model; second, we delve into the interaction of the equations in the model; and third, we look at the output of the FINPLAN model.

Table 1 Warren and Shelton Model

Section 1 – Generation of Sales and Earnings before Interest and Taxes for Period t .

$$(1) SALES_t = SALES_{t-1}(1 + GSALS_t)$$

$$(2) EBIT_t = REBIT_t SALES_t$$

Section 2 – Generation of Total Assets Required for Period t .

$$(3) CA_t = RCA_t SALES_t$$

$$(4) FA_t = RFA_t SALES_t$$

$$(5) A_t = CA_t + FA_t$$

Section 3 – Financing the Desired Level of Assets.

$$(6) CL_t = RCL_t SALES_t$$

$$(7) NF_t = (A_t - CL_t - PFDSK_t) - (L_{t-1} - LR_t) - S_{t-1} - R_{t-1} - b_r \{ (1 - T_t) [EBIT_t - i_{t-1} (L_{t-1} - LR_t)] - PFDIV_t \}$$

$$(8) NF_t + b_t (1 - T_t) [i_t^e NL_t + U^l_t NL_t] = NL_t + NS_t$$

$$(9) L_t = L_{t-1} - LR_t + NL_t$$

$$(10) S_t = S_{t-1} + NS_t$$

$$(11) R_t = R_{t-1} + b_t \{ (1 - T_t) [EBIT_t - i_t L_t - U^l_t NL_t] - PFDIV_t \}$$

$$(12) i_t = i_{t-1} \left(\frac{L_{t-1} - LR_t}{L_t} \right) + i_t^e \frac{NL_t}{L_t}$$

$$(13) \frac{L_t}{S_t + R_t} = K_t$$

Section 4 – Generation of Per Share Data for Period t

$$(14) EAFCD_t = (1 - T_t) [EBIT_t - i_t L_t - U^l_t NL_t] - PFDIV_t$$

$$(15) CMDIV_t = (1 - b_t) EAFCD_t$$

$$(16) NUMCS_t = NUMCS_{t-1} + NEWCS_t$$

$$(17) NEWCS_t = \frac{NS_t}{(1 - U^s_t) P_t}$$

$$(18) P_t = m_t EPS_t$$

$$(19) EPS_t = \frac{EAFCD_t}{NUMCS_t}$$

$$(20) DPS_t = \frac{CMDIV_t}{NUMCS_t}$$

Source: Lee and Lee (2017), page 1166.

There are 20 equation and 20 unknowns for this model. The unknowns are listed and defined in Table 2. In addition, Table 2 lists and defines parameters (inputs)

The inputs to the model are shown in Table 2b. The driving force of the WS model is the sales growth estimates ($GSALS_t$). Equation (1) in Table 1 shows that sales for period t is the product of sales in the prior period multiplied by the growth rate in sales for period t . $EBIT$ is then derived, by expressing $EBIT$ as a percentage of sales ratio, as in Equation (2) of Table 1. Current and fixed assets are then derived in Eqs. 3 and 4 of the table through use of the $CA/SALES$ and $FA/SALES$ ratios. The sum of CA and FA is the total assets for the period.

Financing of the desired level of assets is undertaken in Section 3 of the table. In Eq. 6, current liabilities in period t are derived from the ratio of $CL/SALES$ multiplied by $SALES$. Equation 7 represents the funds required (NF_t). FINPLAN assumes that the amount of preferred stock is constant over the planning horizon. In determining what funds are needed and where they are to come from, FINPLAN uses a source-and-use-of-funds accounting identity. For instance, Eq. 7 shows that the assets for period t are the basis for the firm's financing needs. Current liabilities, as determined in the prior equation, are one source of funds and therefore are subtracted from asset levels. As mentioned above, preferred stock is a constant and therefore must be subtracted also. After the first term in Eq. 7, $(A_t - CL_t - PFDSK_t)$, we have the

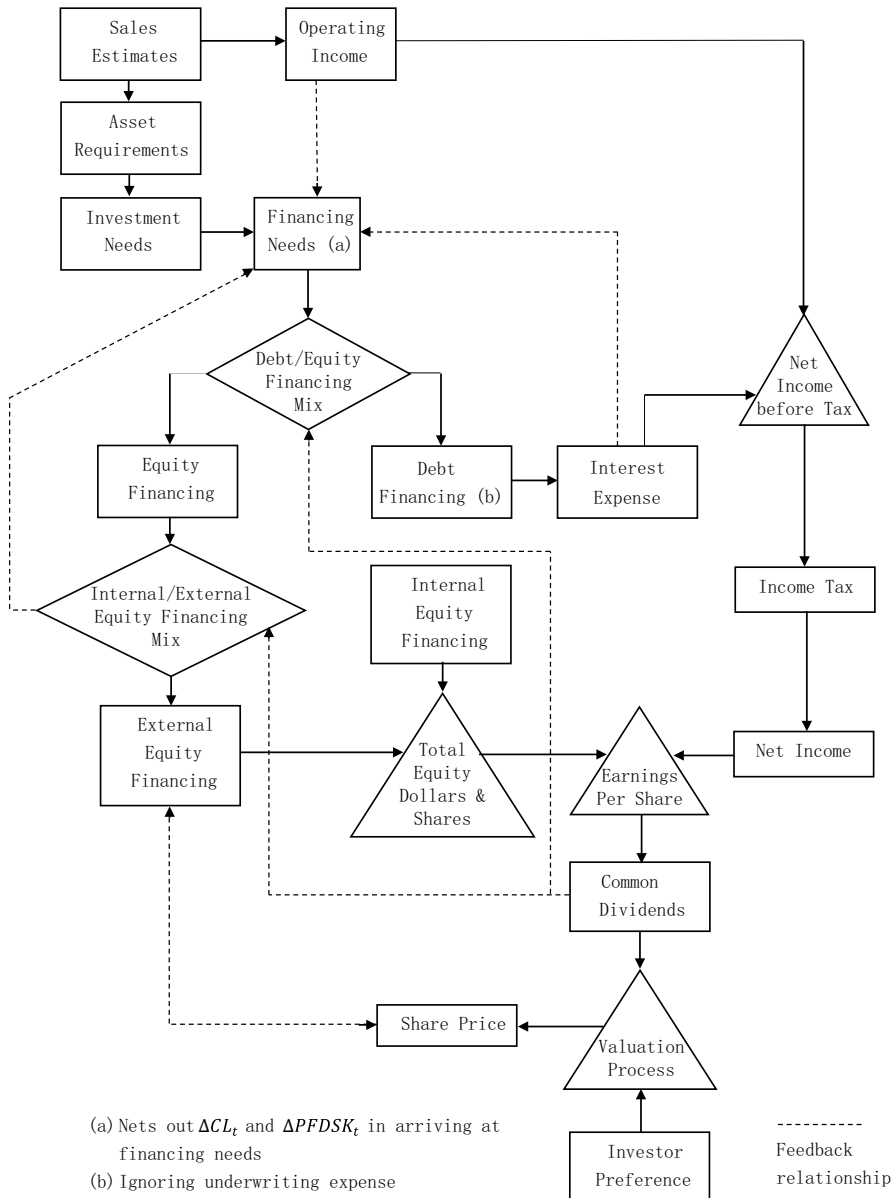


Fig. 2 Flow chart of a simplified financial planning model

financing that must come from internal sources (retained earnings and operations) and long-term external sources (debt and stock issues). The term in the second parenthesis, $(L_{t-1} - LR_t)$, takes into account the remaining old debt outstanding, after retirements, in period t . Then the funds provided by existing stock and retained

Table 2 List of unknowns and list of parameters provided by management

A. Unknowns	
1. $SALES_t$	Sales
2. CA_t	Current assets
3. FA_t	Fixed assets
4. A_t	Total assets
5. CL_t	Current payables
6. NF_t	Needed funds
7. $EBIT_t$	Earnings before interest and taxes
8. NL_t	New debt
9. NS_t	New stock
10. L_t	Total debt
11. S_t	Common stock
12. R_t	Retained earnings
13. I_t	Interest rate on debt
14. $EAFC D_t$	Earnings available for common dividends
15. $CMDIV_t$	Common dividends
16. $NUMCS_t$	Number of common shares outstanding
17. $NEWCS_t$	New common shares issued
18. P_t	Price per share
19. EPS_t	Earnings per share
20. DPS_t	Dividends per share
B. Provided by Management	
21. $SALES_{t-1}$	Sales in previous period
22. $GSALS_t$	Sustainable growth rate
23. RCA_t	Current assets as a percent of sales
24. RFA_t	Fixed assets as a percent of sales
25. RCL_t	Current payables as a percent of sales
26. $PFDSK_t$	Preferred stock
27. $PFDIV_t$	Preferred dividends
28. L_{t-1}	Debt in previous period
29. LR_t	Debt repayment
30. S_{t-1}	Common stock in previous period
31. R_{t-1}	Retained earnings in previous period
32. b_t	Retention rate
33. T_t	Average tax rate
34. i_{t-1}	Average interest rate in previous period
35. i_t^e	Expected interest rate on new debt
36. $REBIT_t$	Operating income as a percent of sales
37. U_t^I	Underwriting cost of debt
38. U_t^S	Underwriting cost of equity
39. K_t	Ratio of debt to equity
40. $NUMSCS_{t-1}$	Number of common shares outstanding in previous period
41. m_t	Price-earnings ratio

Source: Lee and Lee (2017), page 1168

earnings are subtracted out. The last quantity is the funds provided by operations during period t .

Once the funds needed for operations are defined, Eq. 8 specifies that new funds, after taking into account underwriting costs and additional interest costs from new debt, are to come from long-term debt and new stock issues. Equations 9 and 10 simply update the debt and equity accounts for the new issues. Equation 11 updates the retained-earnings account for the portion of earnings available to common stockholders from operations during period t . Specifically, b^t is the retention rate in period t and $(1 - T^t)$ is the after-tax percentage, which is multiplied by the earnings from the period after netting out interest costs on both new and old debt. Since preferred stockholders must be paid before common stockholders, preferred dividends must be subtracted from funds available for common stockholders. Equation 12 calculates the new weighted-average interest rate for the firm's debt. Equation 13 is the new debt-to-equity ratio for period t .

Section 4 of Table 1 applies to the common stockholder; in particular, dividends and market value. Equation 14 represents the earnings available for common dividends and is simply the firm's after-tax earnings. Correspondingly, Eq. 15 computes the earnings to be paid to common stockholders. Equation 16 updates the number of common shares for new issues.

As Eq. 17 shows, the number of new common shares is determined by the total new stock issue divided by the stock price after discounting for issuance costs. Equation 18 determines the price of the stock through the use of a price-earnings ratio (m^t) of the stock purchase. Equation 19 determines *EPS*, as usual, by dividing earnings available to common stockholders by the number of common shares outstanding. Equation 20 determines dividends in a similar manner.

Tables 3, 4, and 5 illustrate the setup of the necessary input variables and the resulting output of the pro forma balance sheet and income statement for the Exxon Company. As mentioned, the WS equation system requires values for parameter inputs, which for this example are listed in Table 3. The first column represents the value of the input, while the second column corresponds to the variable number. The third and fourth columns pertain to beginning and ending periods for the desired planning horizon.

From Tables 4 and 5 you can see the type of information the FINPLAN model generates. With 2016 as a base year, planning information is forecasted for the firm over the period 2017–2020. Based on the model's construction, its underlying assumptions, and the input data, the WS model reveals the following:

1. The amount of investment to be carried out
2. How this investment is to be financed
3. The amount of dividends to be paid
4. How alternative policies can affect the firm's market value

Even more important, as we will explore later in this chapter, this model's greatest value (particularly for FINPLAN) arises from the sensitivity analysis that can be

Table 3 FINPLAN inputs

	Value of data	Variable number	Beginning period	Last period
The number of years to be simulated	4	1	0	0
Net sales at $t-1=2018$	81,581	2	0	0
Sustainable growth rate	0.1076	3	1	4
Current assets as a percent of sales	0.5643	4	1	4
Fixed assets (total – current) as a percent of sales	1.3106	5	1	4
Current payables as a percent of sales	0.3828	6	1	4
Preferred stock	0	7	1	4
Preferred dividends	0	8	1	4
Long-term debt in 2018	27,684	9	0	0
Long-term debt repayment (reduction)	1,555	10	1	4
Common stock in 2018	3,120	11	0	0
Retained earnings in 2018	106,216	12	0	0
Retention rate= $1-\text{payout ratio}$ Payout ratio= $\text{Dividend per share}/$ $\text{earnings per share}$	0.3794	13	1	4
Average tax rate (income taxes/pretax income)	0.1501	14	1	4
Average interest rate in 2018	0.0363	15	0	0
Expected interest rate on new debt	0.0368	16	1	4
Operating income as a percentage of sales	0.2329	17	1	4
Underwriting cost of debt	0.02	18	1	4
Underwriting cost of equity	0.01	19	1	4
Ratio of long-term debt to equity	1.5598	20	1	4
Number of common shares outstanding	2,681.50	21	0	0
Price-earnings ratio	22.3099	22	1	4

performed.² That is, by varying one or several of the input parameters, the financial manager can better understand how his or her decisions interact and, consequently, how they will affect the company's future. (Sensitivity analysis is discussed in greater detail later in this chapter.)

We have shown how we can use Excel to solve 20 simultaneous equation systems as presented in Table 1, and the results are presented in Table 4 and Table 5. Now, we

²The Excel version of the FINPLAN program can be found in Chap. 24 of Lee and Lee (2017). It might be useful for the reader to go through the FINPLAN program by themselves.

Table 4 Pro forma balance sheet (2018–2022)

	2018	2019	2020	2021	2022
Assets					
Current assets	0.00	50984.99	56469.69	62544.40	69272.60
Fixed assets	0.00	118423.00	131162.32	145272.08	160899.69
Total assets	0.00	169407.99	187632.01	207816.48	230172.28
Liabilities and net worth					
Current liabilities	0.00	34589.56	38310.52	42431.77	46996.35
Long-term debt	27684.00	82151.15	90988.54	100776.61	111617.63
Preferred stock	0.00	0.00	0.00	0.00	0.00
Common stock	3120.00	−59003.12	−59711.16	−60495.93	−61365.70
Retained earnings	106216.00	111670.40	118044.10	125104.03	132924.00
Total liabilities and net worth	0.00	169407.99	187632.01	207816.48	230172.28
Computed DBT/EQ	0.0000	1.5598	1.5598	1.5598	1.5598
Int. rate on total debt	0.0363	0.0366	0.0367	0.0367	0.0367
Per share data					
Earnings	0.0000	6.4109	7.5057	8.3296	9.2440
Dividends	0.0000	3.9789	4.6584	5.1697	5.7372
Price	0.0000	143.0258	167.4510	185.8335	206.2330

Table 5 Pro forma income statement (2018–2022)

	2018	2019	2020	2021	2022
Sales	81581.00	90357.06	100077.19	110842.97	122766.88
Operating income	0.00	21048.35	23312.62	25820.47	28598.10
Interest expense	0.00	3010.16	3335.63	3696.05	4095.19
Underwriting commission – debt	0.00	1120.44	207.85	226.86	247.92
Income before taxes	0.00	16917.74	19769.15	21897.56	24254.99
Taxes	0.00	2539.68	2967.73	3287.25	3641.15
Net income	0.00	14378.06	16801.41	18610.31	20613.85
Preferred dividends	0.00	0.00	0.00	0.00	0.00
Available for common dividends	0.00	14378.06	16801.41	18610.31	20613.85
Common dividends	0.00	8923.67	10427.71	11550.39	12793.87
Debt repayments	0.00	1555.00	1555.00	1555.00	1555.00
Actl funds needed for investment	0.00	−6100.97	9684.35	10558.30	11526.24

will discuss how we can use the data from Table 3 to calculate by hand the unknown variables in terms of Johnson & Johnson's 2018 annual report.

Section 1: Generation of Sales and Earnings before Interest and Taxes for Period t

$$\begin{aligned}
 (1) \quad Sales_t &= Sales_{t-1} \times (1 + GSALS_t) \\
 &= 81,581 \times 1.1076 \\
 &= 90,359.1156 \\
 (2) \quad EBIT_t &= REBIT_{t-1} \times Sales_t \\
 &= 0.2329 \times 90,359.1156 \\
 &= 21,044.64
 \end{aligned}$$

Section 2: Generation of Total Assets Required for Period t

$$\begin{aligned}
 (3) \quad CA_t &= RCA_{t-1} \times Sales_t \\
 &= 0.5643 \times 90,359.1156 \\
 &= 50,989.65 \\
 (4) \quad FA_t &= RFA_{t-1} \times Sales_t \\
 &= 1.3106 \times 90,359.1156 \\
 &= 118,424.66 \\
 (5) \quad A_t &= CA_t + FA_t \\
 &= 50,989.65 + 118,424.66 \\
 &= 169,414.31
 \end{aligned}$$

Section 3: Financing the Desired Level of Assets

$$\begin{aligned}
 (6) \quad CL_t &= RCL_{t-1} \times Sales_t \\
 &= 0.3828 \times 90,359.1156 \\
 &= 34,589.47 \\
 (7) \quad NF_t &= (A_t - CL_t - PFDSK_t) - (L_{t-1} - LR_t) - S_{t-1} - R_{t-1} - b_t \times \{(1 - T_t) \\
 &\quad [EBIT_t - i_{t-1}(L_{t-1} - LR_t)] - PFDIV_t\} \\
 &= (169,414.31 - 34,589.47 - 0) - (27,684 - 1,555) - 3,120.0 - 106,216 - \\
 &\quad 0.3794 \times \{(1 - 0.1501) \times [21,044.64 - 0.0363 \times (27,684 - 1,555)] - 0\} \\
 &= -7,120.21
 \end{aligned}$$

Please note that we solve the first seven equations step-by-step; however, Eq. 8 through Eq. 13 are interrelated. Therefore, we need to solve these equations in terms of simultaneous equation system approach. To solve this six-equation system, we first start with Eq. 12 as follows. We substitute parameters into Equation 12 and obtain the following result.

$$\begin{aligned}
 (12) \quad i_t L_t &= i_{t-1}(L_{t-1} - LR_t) + i^e_{t-1} \times NL_t \\
 &= 0.0363 \times (27,684 - 1,555) + 0.0363 \times NL_t \\
 &= 948.4827 + 0.0363 \times NL_t
 \end{aligned}$$

We then substitute related parameters into Equations 8, 9, 10, 11, and 13 to create new a simultaneous equation system to solve unknowns NL_t , NS_t , L_t , S_t , R_t , and i_t .

$$(8) \quad NF_t + b_t(1 - T)[i_{t-1} \times NL_t + U^L_t \times NL_t] = NL_t + NS_t$$

$$-7,120.21 + 0.3794 \times (1 - 0.1501) \times (0.0363 \times NL_t + 0.02 \times NL_t) = NL_t + NS_t$$

$$-7,120.21 + 0.01815 \times NL_t = NL_t + NS_t$$

$$(a) \quad NS_t + 0.98185NL_t = -7,120.21$$

$$(9) \quad L_t = L_{t-1} - LR_t + NL_t$$

$$(b) \quad L_t = 27,684 - 1,555 + NL_t$$

$$L_t - NL_t = 26,129$$

$$(10) \quad S_t = S_{t-1} + NS_t$$

$$(c) \quad -NS_t + S_t = 3,120.0$$

$$(11) \quad R_t = R_{t-1} + b_t\{(1 - T_t)[EBIT_t - i_tL_t - U^L_tNL_t] - PFDIV_t\}$$

$$= 106,216 + 0.3794 \times \{(1 - 0.1501) \times [21,044.64 - i_tL_t - 0.02 \times NL_t] - 0\}$$

Substitute (12) into (11)

$$6480.047315952 - 0.018154050978 \quad NL_t$$

$$R_t = 110,551 + 0.3794 \times \{0.8499 \times [21,044.64 - (948.4827$$

$$+ 0.0363 \times NL_t) - 0.02 \times NL_t]\}$$

$$= 117,031.04731 - 0.01815NL_t$$

$$(d) \quad 117,031.04731 = R_t + 0.01815NL_t$$

$$(13) \quad L_t = (S_t + R_t)K_t$$

$$L_t = 1.5598S_t + 1.5598R_t$$

$$(e) \quad L_t - 1.5598S_t - 1.5598R_t = 0$$

$$(b) - (e) = (f)$$

$$26,129 = 1.5598S_t + 1.5598R_t - NL_t$$

$$(f) - 1.5598(c) = (g)$$

$$21,262.424 = 1.5598NS_t - NL_t + 1.5598R_t$$

$$(g) - 1.5598(d) = (h)$$

$$-NL_t + 1.5598NS_t - 0.0283NL_t = -161,282.6036$$

$$-161,282.6036 = 1.5598NS_t - 1.0283NL_t$$

$$(h) - 1.5598(a) = (i)$$

$$-1.0283NL_t - 1.5315NL_t = -150,176.5$$

$$NL_t = 150,176.5/2.5598 = 58667.279$$

Substitute NL_t in (a)

$$NS_t = -64,722.677$$

Substitute NL_t in (b)

$$L_t = 84,796.279$$

Substitute NS_t in (c)

$$S_t = -61,602.677$$

Substitute NL_t in (d)

$$R_t = 118,095.8584$$

Substitute NL_tL_t in (12)

$$i_t = 0.0363$$

Section 4: Generation of Per Share Data for Period t

$$\begin{aligned} (14) \quad EAFCD_t &= (1 - T_t)(EBIT_t - i_tL_t - U^L_tNL_t) - PFDIV_t \\ &= (1 - 0.1501) \times [21,044.64 - 0.0363 \times 84,796.279 - 0.02 \times (58667.279)] - 0 \\ &= 14,272.532 \end{aligned}$$

$$\begin{aligned} (15) \quad CMDIV_t &= (1 - b_t)EAFCD_t \\ &= (1 - 0.3794) \times (14,272.532) \\ &= 8,857.533 \end{aligned}$$

Please note that we solve Eqs. 14 and 15 step-by-step; however, there are simultaneous equation relationships among Eqs. 16, 17, 18, 19, and 20. Now we show how this five-equation system can be used to solve $NUMCS_t$, $NEWCS_t$, P_t , EPS_t , and DPS_t .

$$(16) \quad NUMCS_t = X_1 = NUMCS_{t-1} + NEWCS_t$$

$$X_1 = 2,681.50 + NEWCS_t$$

$$(17) \quad NEWCS_t = X_2 = NS_t/(1 - U^E_t)P_t$$

$$X_2 = -64,722.677/(1 - 0.01)P_t$$

$$(18) \quad P_t = X_3 = m_tEPS_t$$

$$X_3 = 22.3099(EPS_t)$$

$$(19) \quad EPS_t = X_4 = EAFCD_t / NUMCS_t$$

$$X_4 = 14,272.532 / NUMCS_t$$

$$(20) \quad DPS_t = X_5 = CMDIV_t / NUMCS_t$$

$$X_5 = 8,857.533 / NUMCS_t$$

$$(A) = \text{For (18) and (19), we obtain } X_3 = 22.3099(14,272.532 / NUMCS_t) \\ = 318,418.76166 / X_1$$

Substitute (A) into Eq. (17) to calculate (B)

$$(B) = X_2 = -64,722.677 / [(1 - 0.01) \times 318,418.76166 / X_1]$$

$$(B) = X_2 = -0.2053X_1$$

Substitute (B) into Eq. (16) to calculate (C)

$$(C) = X_1 = 2,681.50 - 0.2053X_1$$

$$(C) = X_1 = 2224.7573 = NUMCS_t$$

Substitute (C) into (B)

$$(B) = X_2 = -456.74 = NEWCS_t$$

From Eqs. (19) and (20) we obtain X_4, X_5

$$X_4 = 6.42 = EPS_t$$

$$X_5 = 3.98 = DPS_t$$

From Eq. (18) we obtain X_3

$$X_3 = 143.2296 = P_t$$

Now we summarize the forecasted variables for 2019 as follows:

- Sales = \$90,359.1156
- Current Assets = \$50,989.65
- Fixed Assets = \$118,424.66
- Total Assets = \$169,414.31
- Current Payables = \$34,589.47
- Needed Funds = (\$7,120.21)
- Earnings before Interest and Taxes = \$21,044.64
- New Debt = \$58,667.279
- New Stock = (\$64,722.68)
- Total Debt = \$84,796.279

Table 6 Mean absolute deviation errors for comparing actual vs. forecasted numbers for 2019

	2019 Forecast	2019 Actual	Mean absolute deviation error for forecast variables
Sales	90,357.06	82,059	10.11%
Operating income	21,048.35	23,416	10.11%
Interest expense	3,010.16	318	846.59%
Income before taxes	16,917.74	17,328	2.37%
Taxes	2,539.68	2,209	14.97%
Net income	14,378.06	15,119	4.90%
Preferred dividends	0	0	0.00%
Common dividends	14,378.06	9,917	44.98%
Debt repayments	1,555	2,823	44.92%
Assets			
Current assets	50,984.99	45,274	12.61%
Fixed assets	118,423	17,658	570.65%
Total assets	169,407.99	157,728	7.41%
Liabilities and net worth			
Current liabilities	34,589.56	35,964	3.82%
Long-term debt	82,151.15	26,494	210.07%
Preferred stock	0	0	0.00%
Common stock	-59,003.12	3,120	1991.13%
Retained earnings	111,670.4	110,659	0.91%
Total liabilities and net worth	169,407.99	176,237	3.87%
Computed DBT/EQ	1.56	1.65	5.58%
Int. rate on total debt	0.04	0.01	233.3%
Per share data			
Earnings	6.4109	5.72	12.08%
Dividends	3.9789	3.75	-6.10%
Price	143.0258	145.75	1.87%

- Common Stock = (\$61,602.677)
- Retained Earnings = \$118,095.86
- Interest Rate on Debt = 3.63%
- Earnings Available for Common Dividends = \$14,272.532
- Common Dividends = \$8,857.53
- Number of Common Shares Outstanding = 2,224.76
- New Common Shares Issued = (456.74)
- Price per Share = \$143.23
- Earnings per Share = \$6.42
- Dividends per Share = \$3.98

The above forecasted variables are almost identical to the numbers for 2019 presented in Tables 4 and 5. We use mean absolute deviation error criteria to

compare these forecasted numbers with actual numbers, and the results are presented in Table 6.

99.5 Sensitivity Analysis

The efficiency of the financial planning and forecasting model will depend solely on how it is used. This section looks at alternative uses of financial planning models to improve their information dissemination. One of the most advantageous ways to use these financial planning models is to perform *sensitivity analysis*. The purpose of sensitivity analysis is to hold all but one or perhaps a couple of variables constant and then analyze the impact of their change on the predicted outcome.

Financial planning models are merely forecasting tools to help the financial manager analyze the interactions of important company decisions with uncertain economic elements. Since we can never be precisely sure what the future holds, sensitivity analysis stands out as a desirable manner of examining the impact of the unexpected as well as of the expected.

The simultaneous equations approach, as embodied in Warren and Shelton’s FINPLAN, offers the best method for performing sensitivity analysis. By changing the parameter values, we can compare new outputs of the financial statements with those such as in Tables 4 and 5. The difference between the new statement and the statements in Tables 4 and 5 reflects the impact of potential changes in such areas as *economic conditions* (reflected in the interest rate, tax rate, and sales growth estimates) and *company policy decisions* (reflected in the maximum and minimum limits specified for the maturity and amount of debt and in the dividend policy as reflected in the specified payout ratio).

To perform sensitivity analysis, we change growth in sales (variable 3), operating income as a percentage of sales (variable 17), the P/E ratio (variable 22), the expected interest rate on new debt (variable 16), and long-term debt-to-equity ratio (variable 20). The new parameters are listed in Table 7. Summary results of the alternative sensitivity analyses for *EPS*, *DPS* and price per share (*PPS*) are listed in

Table 7 Sensitivity analysis parameters

Model variable number	Parameter	Alternative values	Sensitivity analysis number
3	Growth in sales	.20	1
		.15	2
20	Long-term debt-to-equity ratio	.10	3
		.5	4
17	Operating income as a percentage of sales	.20	5
		.50	6
22	Price-to-earnings ratio	5	7
		30	8
16	Expected interest rate on new debt	.005	9
		.05	10

Table 8. The results indicate that changes of key financial decision variables will generally affect *EPS*, *DPS*, and *PPS*.

99.6 Francis and Rowell Model

Francis and Rowell (1978) extended Warren and Shelton's (1971) model from 4 sectors into 10 sectors. The object of their model is to generate *pro forma* financial statements that describe the future financial condition of the firm for any assumed pattern of sales. Parameters of various equations in the system can be changed to answer "what if" questions, perform sensitivity analysis, and explore various paths toward some goals or goals that may or may not be optimal. The variable used by the FR model is presented in Table 9. The completed equation system is presented in Table 10.

The model incorporates an explicit treatment of risk by allowing for stochastic variability in industry sales forecasts. The exogenous input of sales variance is transformed (through simplified linear relations in the model) to coefficients of variation for *EBIT* and net income after taxes (*NIAT*). These are used in risk–return functions that determine the costs of new financing.

Lee and Rahman (1997) use dynamic optimal control model to discuss the interactions of investment, financing, and dividend decisions. Lee and Rahman's approach can integrate with Francis and Rowell's model.

The model also incorporates some variables external to the firm that are important from a financial-planning viewpoint. These industry or economy-wide variables are introduced in every sector to enable the financial planner to explore their influence on plans. They include: market share, an industry capacity-utilization index, the tax rate, and a GNP component price index for explicit analysis of the effects of inflation.

The *FR* model explicitly allows for divergence between planned (or potential) and actual levels in both sales and production. That is, sales forecasts and production potential are compared to determine the existence of slack or idle capacity and company expansion possibilities. Any positive difference between potential or forecasted company sales and actual company sales is decomposed into the portion facilities. As a result, a forecasted sales increase need not lead to investment in new capital. Likewise, a forecasted sales downturn would not lead to a divestiture of capital. An advantage of this disaggregation is that it allows for greater realism – that it, it permits both a lagged production response to sales upturns and downturns, as well as lags, overadjustment, and under adjustment in new investment decisions.

The *FR* model offers a disaggregation of the sales equation into separate market share, production, and pricing equations, which has several distinct advantages. It offers the opportunity to treat sales forecast in physical units that can be compared to technical production capabilities in physical units for both potential and actual levels of sales and production. Such disaggregation also allows distinction between physical units of sales and production and dollar units. Therefore, the pricing decision can

Table 8 Summary results of sensitivity analysis for EPS, DPS, and PPS (2017–2020)

Original analysis	2017	2018	2019	2020
<i>EPS</i>	6.73	7.18	7.63	8.10
<i>DPS</i>	3.51	3.74	3.97	4.22
<i>PPS</i>	128.29	136.96	145.45	154.48
Sensitivity analysis #1				
<i>EPS</i>	7.35	8.66	10.15	11.91
<i>DPS</i>	3.83	4.51	5.29	6.21
<i>PPS</i>	140.23	165.17	193.68	227.12
Sensitivity analysis #2				
<i>EPS</i>	5.89	5.40	4.94	4.52
<i>DPS</i>	3.07	2.81	2.58	2.36
<i>PPS</i>	112.38	103.00	94.25	86.23
Sensitivity analysis #3				
<i>EPS</i>	6.90	7.71	8.58	9.55
<i>DPS</i>	3.60	4.02	4.47	4.98
<i>PPS</i>	131.58	147.03	163.62	182.10
Sensitivity analysis #4				
<i>EPS</i>	7.07	8.05	9.03	10.14
<i>DPS</i>	3.68	4.20	4.71	5.29
<i>PPS</i>	134.82	153.56	172.34	193.42
Sensitivity analysis #5				
<i>EPS</i>	4.88	5.41	5.96	6.56
<i>DPS</i>	2.54	2.82	3.10	3.42
<i>PPS</i>	93.01	103.17	113.62	125.14
Sensitivity analysis #6				
<i>EPS</i>	12.34	14.04	15.93	18.08
<i>DPS</i>	6.43	7.32	8.30	9.42
<i>PPS</i>	235.39	267.81	303.88	344.82
Sensitivity analysis #7				
<i>EPS</i>	8.36	9.22	10.11	8.36
<i>DPS</i>	4.36	4.80	5.27	4.36
<i>PPS</i>	41.82	46.08	50.53	41.82
Sensitivity analysis #8				
<i>EPS</i>	6.88	7.75	8.69	9.74
<i>DPS</i>	3.58	4.04	4.53	5.08
<i>PPS</i>	206.26	232.42	260.65	292.32
Sensitivity analysis #9				
<i>EPS</i>	7.15	8.09	9.09	10.21
<i>DPS</i>	3.73	4.22	4.74	5.32
<i>PPS</i>	136.48	154.27	173.38	194.74
Sensitivity analysis #10				
<i>EPS</i>	7.00	7.85	8.76	9.79
<i>DPS</i>	3.65	4.09	4.57	5.10
<i>PPS</i>	133.53	149.73	167.15	186.65

EPS Earning per share, *DPS* Dividend per share, *PPS* price per share

Table 9 List of variables for FR model

<i>Endogenous</i>		<i>Exogenous</i>	
$Sales_t^P$	Potential industry sales (units)	$GSALS_t$	Growth rate in potential industry sales
S_t^{FC}	Full capacity unit output (company)	$Sales_{t-1}^P$	Previous period potential industry sales (units)
S_t^a	Actual company unit output	S_{t-1}^{FC}	Previous period company full capacity unit output
S_t^P	Potential company unit output	INV_{t-1}	Previous period company finished goods inventory
γ_{1t}	Measure of necessary new investment (based on units)	FA_{t-1}	Previous period company fixed asset base (\$)
γ_{2t}	Measure of slack due to underutilization of existing resources	γ_t	Capacity utilization index
K_t	Units of capital stock	c_t	Desire market share
NK_t	Desired new capital (capital units)	θ	Proportionality coefficient of S_t^{RC} to K_t
FA_t	Fixed assets (current \$)	P_{kt}	GNP component index for capital equipment
NF_t	Desired new investment (current \$)	P	Percentage markup of output price over ratio of GOP_t/INV_t
P_{ts}	Output price	δ_2	Proportionality coefficient of OC_t to $\$S_t$
$\$S_t$	Sales dollars (current \$)	Φ	Proportionality coefficient of D_t to FA_t
COG_t	Cost of goods (current \$)	N	Proportionality coefficient of INV_t to $\$S_t$
OC_t	Overhead, selling, cost of goods (current \$)	LR_t	Repayment of long-term debt
$OC2_t$	Nonoperating income (current \$)	T_t	Corporate tax rate
D_t	Depreciation expense (current \$)	b_t	Retention rate
INV_t	Inventory (current \$)	U_t^L	Underwriting cost of new debt
L_t	Long-term debt	$PFDIV_t$	Preferred dividend
i_t^L	Cost of new debt (%)	i_{t-1}^A	Previous period weighted average cost of long-term debt
NL_t	New long-term debt needed (\$)	L_{t-1}	Previous period long-term debt
NS_t	New common stock (equity) needed (\$)	k	Optimal capital structure assumption
$NIAT_t$	Net income after tax (current \$)	α_L, β_L	Coefficients in risk-return tradeoff for new debt
RE_t	Retained earnings	α_s, β_s	Coefficients in risk-return tradeoff for new stock
$EBIT_t$	Earnings before interest and taxes	GOP_{t-1}	Gross operating profit of previous period
i_t^A	Weighted average cost of long term debt	δ_1	Ratio of COG_t to actual net sales

(continued)

Table 9 (continued)

<i>Endogenous</i>		<i>Exogenous</i>	
v_{EBIT}	Coefficient of variation of EBIT	δ_3	Ratio of OC2 to net sales
i_t^s	Cost of new stock issue	$\alpha_1, \alpha_2, \alpha_3$	Production function coefficients
v_{NIAT}	Coefficient of variation of NIAT	$\sum_1$	Ratio of CA_t to net sales
TEV_t	Total equity value	$\sum_2$	Ratio of CL_t to net sales
g_t^a	Growth rate in $\$S_t$	$\sigma_{Sales}^2 \rho$	Standard deviation of potential industry sales
$EAFCD_t$	Earnings available for common dividend		
$CMDIV_t$	Common dividend		
ΔRE_t	Contributions to RE made in the t^{th} period		
GPO_t	Gross operating profit (current \$)		

Source: Lee and Lee (2017), pages 1188–1189

Table 10 List of Equations for FR Model

1. <i>Industry sales</i> (1) $Sales_t^p = Sales_{t-1}^p (1 + GSALS_t)$ 2. <i>Company production sector</i> (2) $S_t^{FC} = \alpha_1 S_{t-1}^{FC} + \alpha_2 INV_{t-1} + \alpha_3 FA_{t-1}$ (3) $\frac{S_t^a}{S_t^{FC}} = \gamma_t \rightarrow S_t^a = \gamma_t S_t^{FC}$ (4) $S_t^p = c_t Sales_t^p$ 3. <i>Capital stock requirements sector</i> (5) $S_t^p - S_t^a = (S_t^{FC} - S_t^a) + (S_t^p - S_t^{FC})$ (6) $S_t^{FC} - S_t^a = \gamma_{2t}$ (7) $S_t^p - S_t^{FC} = \gamma_{1t} \quad (0 \leq \gamma_{1t})$ (8) $K_1 = \theta S_t^{FC}$ (9) $NK_t = \theta \gamma_{1t}$ 4. <i>Pricing sector</i> (10) $P_{Kt} \cdot gK_t = FA_t$ or $FA_t/K_t = P_{Kt}$ (11) $P_{Kt} \cdot gNK_t = NF_t$ (12) $P_{st} \cdot gS_t^a = \\$S_t^a$ (13) $P_{is} = p(GOP_{t-1}/INV_{t-1})$ 5. <i>Production cost sector</i> (14) $OC_t = \delta_2 (\\$S_t^a)$ (15) $COG_t = \delta_1 (\\$S_t^a)$ (16) $GOP_t = \\$S_t^a - COG_t$ (17) $OC2_t = \delta_3 (\\$S_t^a)$ 6. <i>Income sector</i> (18) $INV_t = N(\\$S_t^a)$ (19) $EBIT_t = \\$S_t^a - OC_t + OC2_t - D_t$ (20) $NIAT_t = (EBIT_t - i_t^A L_t)(1 - T)$ (20') $CL_t = \sum_2 (\\$S_t^a)$	7. <i>New financing required sector</i> (21) $NF_t + b_t \{ (1 - T) [i_t^A NL_t + U_t^L NL_t] \}$ $= NLS_t + \Delta RE_t + (CL_t - CL_{t-1})$ (22) $NLS_t = NS_t + NL_t$ (23) $\Delta RE_t = b_t \{ (1 - T) [EBIT_t - i_t^A L_t - U_t^L NL_t] - PFDIV_t \}$ (24) $i_t^A = i_{t-1}^A \left[\frac{L_{t-1} - LR_t}{L_t} \right] + i_t^L \frac{NL_t}{L_t}$ (25) $\frac{NL_t}{NS_t + \Delta RE_t} = k$ (26) $L_t = L_{t-1} - LR_t + NL_t$ 8. <i>Risk sector</i> (27) $\sigma_{ebit}^2 = \theta_1^2 \cdot g\theta_2^2 \cdot g\sigma_{Sales_t^p}^2$ (28) $\sigma_{niat}^2 = \theta_3^2 \cdot g\theta_6^2 \cdot g\theta_2^2 \cdot g\sigma_{Sales_t^p}^2$ 9. <i>Costs of financing sector</i> (29) $i_t^L = \alpha_L + \beta_L v_{EBIT}$ (30) $v_{EBIT} = \frac{\sigma_{EBIT}}{R_{EBIT}}$ (31) $i_t^a = \alpha_s + \beta_s v_{NIAT}$ (32) $v_{NIAT} = \frac{\sigma_{NIAT}}{R_{NIAT}}$ 10. <i>Valuation of equity sector</i> (33) $TEV_t = \frac{CMDIV_t}{i_t^s - g_t^s}$ (34) $EAFCD_t = (1 - T_t) \times [EBIT_t - i_t^A L_t - U_t^L NL_t] - PFDIV_t$ (35) $CMDIV_t = (1 - b_t) EAFCD_t$ (36) $g_t^a = \frac{\\$S_t - \\$S_{t-1}}{\\$S_{t-1}}$
---	---

Source: Lee and Lee (2017), page 1190

be treated separately. This feature is helpful in analyzing the effect of changing prices.

Another aspect of the *FR* financial-forecasting model is the econometrical advantage. The *FR* model's risk–return function and its production function are estimated econometrically. Additionally, standard econometric techniques to evaluate goodness-of-fit and predictive power of a simultaneous equation system are reported. In the remaining subsections of this section, the *FR* model is explained in its general form. Then the coefficients of the equations are set to equal to the values that characterize the operations of an existing company, and then the active operations of a well-known firm are simulated, to test this financial model empirically.

99.6.1 The FR Model Specification

The *FR* model is composed of 10 sectors: (1) industry sales, (2) production sector, (3) fixed capital-stock requirements, (4) pricing, (5) production costs, (6) income, (7) new financing required, (8) risk, (9) costs of financing, and (10) common stock valuation.

Looking more deeply reveals that the *FR* model is, to a large extent (but not entirely), recursive between sectors. All entries of the sector-interdependence table, with the exception of one (between sectors seven and nine), are below the diagonal. It has been structured in this manner for the specific purpose of ease of exposition and computation. The simultaneity of the *FR* model is primarily within each sector's equations.

99.7 Summary

Entry 102 has examined three types of financial planning models available to the financial manager for use in analyzing the interactions of company decisions: the algebraic simultaneous equations model, the linear programming model, and the econometric model. We also have discussed the benefits of sensitivity analysis for determining the impact on the company from changes (expected and unexpected) in economic conditions.

The student should understand the basic functioning of all three models, along with the underlying financial theory. Moreover, it is essential to understand that a financial planning model is an aid or tool to be used in the decision-making process and is not an end in and of itself.

The computer-based financial modeling discussed in this chapter can be performed on either a mainframe computer or a PC. An additional dimension is the development of electronic spreadsheets. These programs simulate the matrix or spreadsheet format used in accounting and financial statements. Their growing acceptance and popularity is due to the ease with which users can make changes

in the spreadsheet. This flexibility greatly facilitates the use of these programs for sensitivity analysis.

Appendix A – Description of Parameter Inputs Used to Forecast Johnson & Johnson’s Financial Statements and Share Price

In our financial planning plan program, there are 20 equations and 20 unknowns. To use this program, we need to input 21 parameters. These 20 unknowns and 21 parameters can be found in Table 21.2.

We use 2016 as the initial reference year and input the 21 parameters, the bulk of which can be obtained or derived from the historical financial statements of JNJ. The first input is $SALE_{t-1}$ (\$71,890), defined as fiscal 2016 net sales and can be obtained from the income statement of JNJ. The second input is $GCALS_{t-1}$. This parameter can be calculated by either the percentage change method: $\frac{Sales_{t-1} - Sales_{t-2}}{Sales_{t-2}} = 2.59\%$ or sustainability growth rate: $\frac{ROE_{t-1} * b_{t-1}}{1 - ROE_{t-1} * b_{t-1}} = 12.7\%$.

The third input is RCA_{t-1} (90.46%), defined as current assets divided by total sales, and the fourth input is RLA_{t-1} (1.0596), defined as total asset minus current asset divided by net sales. The next parameter is RCL_{t-1} (36.57%), defined as current liabilities as a percentage of net sales. The sixth parameter is preferred stock issued (PKV), with a value of 0, as JNJ does not currently have any preferred stock outstanding. The inputs for the aforementioned three parameters are all obtained from the JNJ’s fiscal 2016 balance sheet. The seventh input is JNJ’s preferred stock dividends, and since there is no preferred stock outstanding, it is correspondingly 0. The eighth input is LR_{t-1} (\$22,442), defined as long-term debt, coming from the balance sheet of JNJ for the fiscal year 2016, and the ninth input is LR_{t-1} (\$-2,223), defined as long-term debt retirement, from the 2016 statement of cash flows.

The tenth input is S_{t-1} (\$3,120), which represents common stock issued, and the eleventh input is retained earnings (R_{t-1} =\$110,551). Both of these two variables can be found in the balance sheet for JNJ’s fiscal year 2016. The twelfth input is the retention rate (b_{t-1} = 47.88%), defined as $1 - \frac{Dividend\ payout_{t-1}}{Net\ income_{t-1}}$. The thirteenth input, the average tax rate (T_{t-1}), is assumed to be 15%. The fourteenth input is the weighted average effective interest rate (I_{t-1} = 3.33%), which JNJ provides in its annual report (page 53 of the respective 10-K filing). The fifteenth input is expected interest on new debt (ie_{t-1} = 3.68%), calculated as the average of the weighted average interest rates in previous two periods.

The next input is $REBIT_{t-1}$ (28.71%), defined as operating income as a percentage of sales. However, JNJ does not explicitly list operating income in its income statements. Thus, we defined operating income as JNJ’s earnings before provision for taxes on income, with interest expense added back and interest income subtracted out. We also adjusted for non-recurring expenses and added back other income/losses (related primarily to hedging activities, write-downs, and

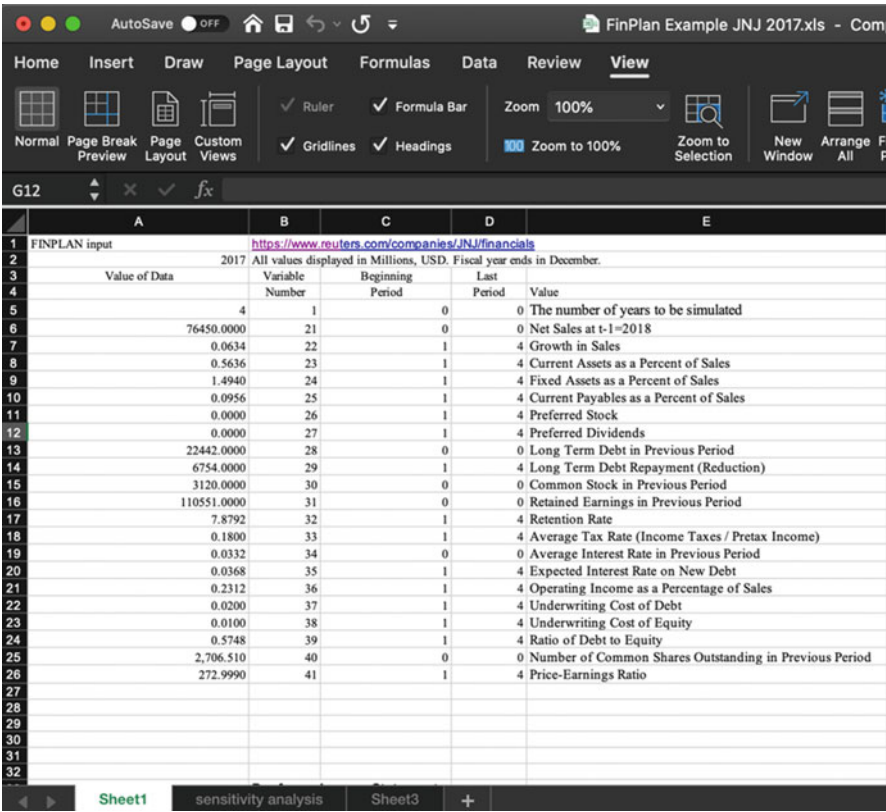
restructuring charges) to get to an adjusted and normalized operating income figure.

The seventeenth input is the underwriting cost of debt (UL) that we assume to be 2%, and the eighteenth parameter is the underwriting cost of equity (UE = 1%). The nineteen input is the ratio of long-term debt to equity ($K_{t-1} = 31.87\%$), defined as long-term debt divided by total equity. The twentieth input is the number of common shares outstanding ($NUMCS_{t-1} = 2,737.3$) listed in the JNJ's Balance Sheet for fiscal year 2016. The last input is P/E ratio ($m_{t-1} = 19.075$), which is calculated as JNJ's closing share price on the last trading day of 2016 divided by fiscal 2016 net income.

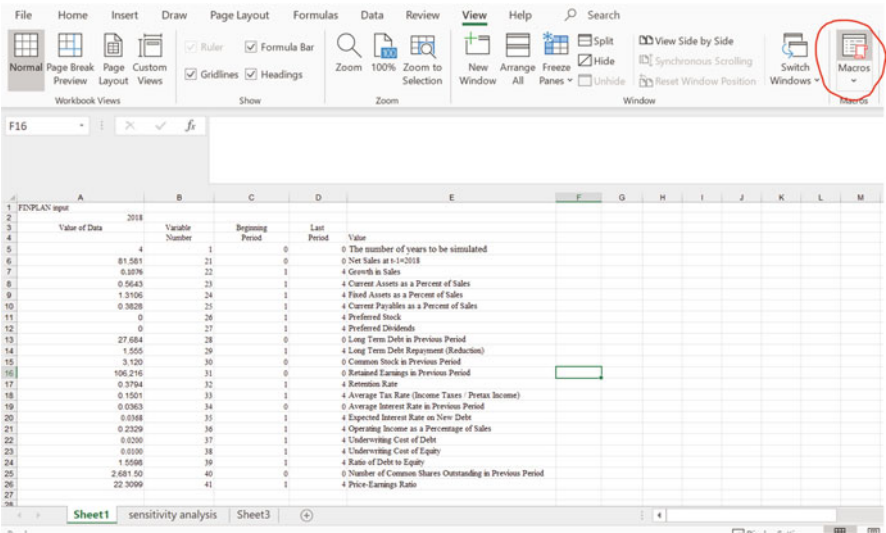
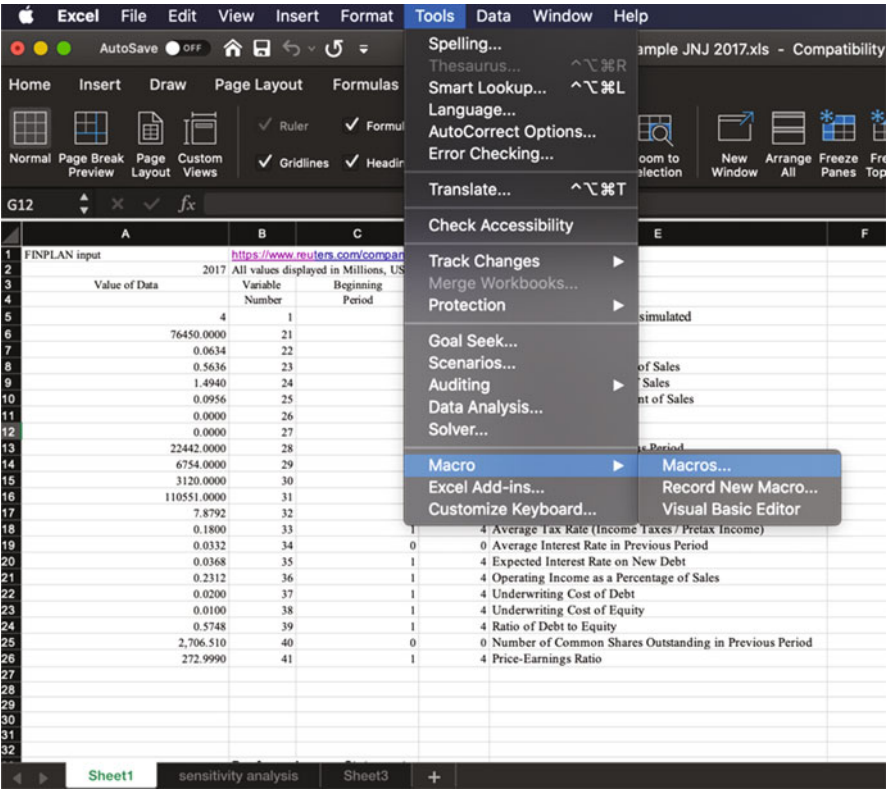
Appendix B – Procedure of Using Excel to Implement the FinPlan Program

This appendix describes the detailed procedure of using Excel to implement the FinPlan program. There are four steps to use the FinPlan program.

Step 1. Open the Excel file of FinPlan Example.



Step 2. Click the “Tools” and see “Macros.” Another way to find it is under “View” and see “Macors.”



Step 3. Choose “Macros” and then click “Run.”

After we obtain the forecasted values from the model, we compare them with the actual data of JNJ in 2019 via calculating the absolute percentage change of error. The following table shows results.

References

- Francis, J.C., and D.R. Rowell. 1978. A Simultaneous Equation Model of the Firm for Financial Analysis and Planning. *Financial Management* 7: 29–44.
- Harrington, D.R. 1985. *Case Studies in Financial Decision-Making*. Chicago: The Dryden Press.
- Lee, C.F., and J. Lee. 2017. *Financial Analysis and Planning: Theory and Application*. 3rd ed. Singapore: World Scientific.
- Lee, C.F., and S. Rahman. 1997. Interaction of Investment, Financing, and Dividend Decisions: A Control Theory Approach. *Advances in Financial Planning and Forecasting* 7: 100–130.
- Myers, S.C. 1974. Interaction of Corporate Financing and Investment Decisions. *Journal of Finance* 29: 1–25.
- Stern, J.M. 1980. The Dynamics of Financial Planning. *Analytical Methods in Financial Planning*: 29–41.
- Taggart, R.A., Jr. 1977. A Model of Corporate Financing Decisions. *Journal of Finance* 32: 1467–1484.
- Warren, J., and J. Shelton. 1971. A Simultaneous Equations Approach to Financial Planning. *Journal of Finance* 26: 1123–1142.
- Ross, S. A., Westerfield, R., Jaffe, J. F., Jordan, B. D., Jaffe, J., & Jordan, B. (2019). Corporate finance (pp. 880–86). McGraw-Hill Education.
- Lee, C. F., Finnerty, J. E., & Norton, E. (1997). *Foundations of financial management*. West Pub.



Alternative Errors-in-Variables Models and Their Applications in Finance Research

100

Hong-Yi Chen, Alice C. Lee, and Cheng-Few Lee

Abstract

Specification error and measurement error are two major issues in finance research. The main purpose of this chapter is (i) to review and extend existing errors-in-variables (EIV) estimation methods, including classical method, grouping method, instrumental variable method, mathematical programming method, maximum likelihood method, LISREL method, and the Bayesian approach; (ii) to investigate how EIV estimation methods have been used to finance related studies, such as cost of capital, capital structure, investment equation, and test capital asset pricing models; and (iii) to give a more detailed explanation of the methods used by Almeida et al. (*Review of Financial Studies*, 23, 3279–3328, 2010).

Keywords

Measurement error · Errors-in-variables · Cost of capital · Capital structure · Investment equation · Capital asset pricing model

This chapter is the revised and updated version of our entry entitled “Alternative Errors-in-Variables Models and their applications in Finance Research” which was published in *The Quarterly Review of Economics and Finance* in 2015.

H.-Y. Chen

Department of Finance, National Chengchi University, Taipei, Taiwan

e-mail: fnhchen@nccu.edu.tw

A. C. Lee (✉)

Center for PBBEF Research, Phoenix, AZ, USA

C.-F. Lee

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: lee@business.rutgers.edu

JEL ClassificationC58 · G10

100.1 Introduction

Specification error and measurement error are two major issues in applying the econometric model to economic and finance research. Studies by Miller and Modigliani (1966) and Roll (1969) are two of the earliest finance-related research studies to apply errors-in-variables (EIV) model in their empirical works. Miller and Modigliani (1966) show that, in determining the cost of capital, anticipated average earnings are unobservable and using accounting estimates of earnings as the proxy may result in measurement error problems. Roll (1969, 1977) and Lee and Jen (1978) show that the observed market rate returns in terms of stock market index are measured with errors since the stock market index does not include all assets that can be invested by investors. Roll (1969, 1977) argues that testing capital asset pricing model suffers from an EIV problem and concludes that no correct and unambiguous test of the theory can be accomplished. Lee and Jen (1978) have theoretically shown how beta estimates and Jensen performance measures can be affected by both constant and random measurement errors of the market rate of return and risk free rate. Other issues such as the determination of the capital structure and investment functions also suffer EIV problems.¹

Understanding the existence of measurement error problems on finance-related studies, a large extent of the literature subsequently tries to mitigate biased results from measurement errors. For the issue of the estimation of the cost of capital, Miller and Modigliani (1966) use the instrumental variable approach to resolve the measurement error problem and get consistent estimators in determining the cost of capital. Zellner (1970) and Lee and Wu (1989) also use various estimation methods to deal with potential EIV problems on estimating the cost of capital. For the issue of the capital asset pricing test, Lee and Jen (1978) argue that both market return and beta coefficient are subjected to measurement error, and show how the beta coefficient can be estimated. Lee (1984) shows that the most generalized beta estimate can be decomposed into three components; bias due to specification error, bias due to measurement error, and interaction bias. Therefore, the evidence of failure in capital asset pricing model or the finding of new risk factors might result from model misspecification error or EIV problem. Gibbons and Ferson (1985), Green (1986), Roll and Ross (1994), and Diacogiannis and Feldman (2011) have argued that market portfolio measure with errors is an inefficient portfolio and show how the

¹For the measurement problems related to the determinants of the capital structure, please see Titman and Wessels (1988), Chang et al. (2009), and Yang et al. (2009). For the measurement problems related to the investment function, please see Erickson and Whited (2000, 2002) and Almeida et al. (2010).

inefficient benchmark can affect the theoretical CAPM derivation. For the issue of the determinants of the capital structure, Titman and Wessel (1998), Chang et al. (2009), and Yang et al. (2009) apply structure equation models to investigate determinants of the capital structure. For the measurement error problems related to Tobin's q in investment function, Erickson and Whited (2000) use generalized method of moments (GMM) to obtain consistent estimators in testing q theory. Most recently, Almeida et al. (2010) propose an alternative instrumental method to deal with measurement error problems in Tobin's q and support the q theory.

The main purpose of this chapter is to study existing EIV estimation methods and to discuss how these estimation methods have been used in finance research. We first show how EIV problems affect estimators in the regression model. We further demonstrate seven alternative estimation methods dealing with EIV problems. Classical method, grouping method, instrumental variable method, mathematical programming method, maxima likelihood method, LISREL method, and Bayesian approach will be discussed. Finally, we conduct a survey on various studies and investigate the effect that resulted from EIV problems associated with cost of capital, capital asset pricing model, capital structure, and investment equation. We also investigate the correction models used in such studies to mitigate the problem raised from measurement errors.

The remainder of this chapter is organized as follows. Section 100.2 shows the classical EIV problems and how they affect estimators of the linear regression model. Section 100.3 provides seven alternative correction methods in dealing with EIV problems. Section 100.4 presents the effects of EIV problems on the empirical research of cost of capital, asset pricing, capital structure, and investment decision. Finally, section 100.5 presents the conclusion.

100.2 Effects of Errors-in-Variables in Different Cases

100.2.1 Bivariate Normal Case

Suppose we have a two variate structural relationship

$$V_i = \alpha + \beta U_i \quad (1)$$

Both V_i and U_i are unobserved, while we can observe $Y_i = V_i + \eta_i$ and $X_i = U_i + \varepsilon_i$. We assume that.

- (a) $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ and $\eta_i \sim N(0, \sigma_\eta^2)$.
- (b) $E(\varepsilon_i U_i) = 0$, $E(\varepsilon_i V_i) = 0$, $E(\varepsilon_i \eta_i) = 0$, $E(\eta_i U_i) = 0$, and $E(\eta_i V_i) = 0$.
- (c) $U_i \sim N(E(X), \sigma_U^2)$ and $V_i \sim N(\alpha + \beta E(X), \beta^2 \sigma_U^2)$.

This results in measurement error on the estimates of α and β implying that the asymptotic biases for β and α are

$$\text{plim} \hat{\beta} - \beta = \frac{-\beta \sigma_1^2}{\sigma_U^2 + \sigma_1^2} \quad (2a)$$

$$\text{plim} \hat{\alpha} - \alpha = \frac{\beta \sigma_1^2}{\sigma_U^2 + \sigma_1^2} E(X). \quad (2b)$$

Equation (2) implies that $\hat{\beta}$ is downward biased and $\hat{\alpha}$ is upward biased.

100.2.2 Multivariate Case

Suppose we have a trivariate structural relationship

$$W_i = \alpha + \beta U_i + \gamma V_i \quad (4)$$

W_i , U_i , and V_i are unobserved, but we can observe $Z_i = W_i + \tau_i$, $X_i = U_i + \varepsilon_i$, and $Y_i = V_i + \eta_i$. U_i and V_i have a joint normal distribution with variances σ_U^2 and σ_V^2 and correlation coefficient ρ_{UV} . In the observed variables X , Y , and Z , the observed errors ε , η , and τ are independent normal variables with zero means and variance σ_ε^2 , σ_η^2 , σ_τ^2 . X , Y , and Z have a multivariate normal distribution.

The asymptotic biases of $\hat{\beta}$ and $\hat{\gamma}$ can be seen from the following:

$$\text{plim} \hat{\beta} - \beta = \frac{\sigma_{VW} \sigma_\eta^2 - \beta (\sigma_U^2 \sigma_\eta^2 + \sigma_V^2 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \sigma_\eta^2)}{(\sigma_U^2 \sigma_V^2 - \sigma_{UV}^2) + \sigma_U^2 \sigma_\eta^2 + \sigma_V^2 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \sigma_\eta^2} \quad (4a)$$

$$\text{plim} \hat{\gamma} - \gamma = \frac{\sigma_{WV} \sigma_\varepsilon^2 - \gamma (\sigma_U^2 \sigma_\eta^2 + \sigma_V^2 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \sigma_\eta^2)}{(\sigma_U^2 \sigma_V^2 - \sigma_{UV}^2) + \sigma_U^2 \sigma_\eta^2 + \sigma_V^2 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \sigma_\eta^2} \quad (4b)$$

The direction of the biases of $\hat{\beta}$ and $\hat{\gamma}$ can be treated according to different assumptions.²

Concerning the coefficient of the reliability, Cochran (1970) shows that measurement errors of both explained and explanatory variables will reduce the multiple correlations and increase the residual variance, and the good prediction formula is more sensitive to measurement errors than the poor one. Moreover, from the analysis of the effects of measurement error on both the simple regression coefficient and residual variance, in general, we can conclude that the t statistic of the simple regression coefficient will be downward biased if variables are measured with errors.

²Please see Lee (1973) and Chen (2011) for detail.

100.3 Estimation Methods when Variables Are Subject to Error

In this section, we will discuss alternative EIV estimation methods, classical method, grouping method, instrumental variable method, mathematical method, maxima likelihood method, LISREL method, and the Bayesian approach.

100.3.1 Classical Estimation Method

100.3.1.1 The Classical Method to a Simple Regression Analysis

In general, the classical method considers three cases: (i) either σ_ε^2 or σ_η^2 is known; (ii) $\lambda = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$ is known; and (iii) σ_ε^2 and σ_η^2 are known. We can obtain the estimate for β from Eq. (2) under every possible situation as:

$$(i) \quad \hat{\beta} = \frac{S_{XY}}{S_{XX} - \sigma_\varepsilon^2}, \text{ when } \sigma_\varepsilon^2 \text{ is known.} \quad (5)$$

$$\hat{\beta} = \frac{S_{YY} - \sigma_\eta^2}{S_{XY}}, \text{ when } \sigma_\eta^2 \text{ is known.} \quad (6)$$

$$(ii) \quad \hat{\beta} = \frac{(S_{YY} - \lambda S_{XX}) + \left\{ (S_{YY} - \lambda S_{XX})^2 + 4\lambda S_{XY} \right\}^{1/2}}{2S_{XY}}, \text{ when } \lambda = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2} \text{ is known.} \quad (7)$$

(iii) When both σ_ε^2 and σ_η^2 are known, Kendall and Stuart (1961) regarded it as an overidentified situation unless a nonzero covariance between U_i and V_i is introduced. Barnett (1967) followed Kiefer's (1964) suggestion and derived a consistent estimator of $\hat{\beta}$ as one of the real roots of Eq. (19).

$$\begin{aligned} & \hat{\beta}^4 - \left(\frac{1}{b_2} - \frac{\lambda}{b_1} - 2\lambda b_2 \right) \hat{\beta}^3 - 3\lambda \left(1 - \lambda \frac{b_2}{b_1} \right) \hat{\beta}^2 + \lambda^2 \left(\lambda \frac{b_2}{b_1^2} - \frac{1}{b_1} - 2b_2 \right) \hat{\beta} \\ & - \lambda^3 \frac{b_2}{b_1} \\ & = 0, \end{aligned} \quad (8)$$

where $S_{XX} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$, $S_{YY} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}$, $S_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}$, $m_{XX} = \text{Var}(X) = \sigma_U^2 + \sigma_\varepsilon^2$, $m_{YY} = \text{Var}(Y) = \beta^2 \sigma_U^2 + \sigma_\eta^2$, $m_{XY} = \text{Cov}(X, Y) = \beta \sigma_U^2$, $b_1 = \frac{m_{XY}}{m_{XX}}$, $b_2 = \frac{m_{XY}}{m_{YY}}$, and $\lambda = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$.

The advantage of the knowledge of both σ_ε^2 and σ_η^2 is that one obtains a more efficient estimator of $\hat{\sigma}_\eta^2$.

The analysis of Eqs. (5), (6), (7), and (8) furnishes us two important implications. First, if only U or V is subject to measurement error, then we know that the maximum likelihood estimator of β is equivalent to fitting a least square line when using the error-free variable as a regressor. Second, if both U and V are measured with errors, then the estimate of β lies between the values estimated from Eqs. (5) and (6). This situation can be further analyzed. The quadratic equation for Eq. (7) is

$$\hat{\beta}^2 S_{XY} + \hat{\beta}(\lambda S_{XX} - S_{YY}) - \lambda S_{XY} = 0. \quad (9)$$

(a) When either $\sigma_\eta^2 \rightarrow 0$ or $\sigma_\varepsilon^2 \rightarrow \infty$, Eq. (6) shows $\hat{\beta} = \frac{S_{YV}}{S_{XV}}$.

(b) When $\sigma_\varepsilon^2 \rightarrow 0$ or $\sigma_\eta^2 \rightarrow \infty$, Eq. (5) shows $\hat{\beta} = \frac{S_{XV}}{S_{XX}}$.

100.3.1.2 The Classical Method to a Multiple Regression Analysis

It is clear that U is orthogonal to V if $\rho = 0$. It is well known that this multiple regression reduces to two simple regression relationships. If $\rho \neq 0$, to identify β and γ , we need to know either the actual values of σ_ε^2 , σ_η^2 , and σ_τ^2 or the relative ratios among σ_ε^2 , σ_η^2 , and σ_τ^2 . We will investigate the following cases:

- (i) $\sigma_\eta^2 = 0$, $\sigma_\varepsilon^2 > 0$, $\sigma_\tau^2 < 0$, and $\sigma_\varepsilon^2 = \lambda \sigma_\tau^2$
- (ii) $\sigma_\varepsilon^2 > 0$, $\sigma_\eta^2 > 0$, and $\sigma_\tau^2 > 0$
 - (a) σ_ε^2 and σ_η^2 are known
 - (b) $\sigma_\eta^2 = \lambda_1 \sigma_\varepsilon^2$ and $\sigma_\tau^2 = \lambda_2 \sigma_\varepsilon^2$

(Case i) Only Z and X are measured with errors.

The estimator of $\hat{\beta}$ can be one of real root of Eq. (10)

$$k_2 \hat{\beta}^2 + k_1 \hat{\beta} + k_0 = 0, \quad (10)$$

where

$$\begin{aligned} k_0 &= \left(S_{XZ} - \frac{S_{XY}S_{YZ}}{S_{YY}} \right), k_1 = - \left(S_{XX} - \frac{S_{XY}^2}{S_{YY}} \right) + \lambda \left(S_{ZZ} - \frac{S_{YZ}^2}{S_{YY}} \right), \text{ and } k_2 = -\lambda k_0 \\ &= -\lambda \left(S_{XZ} - \frac{S_{XY}S_{YZ}}{S_{YY}} \right). \end{aligned}$$

There are three cases to consider:

(a) $\lambda \rightarrow 0$ when $\sigma_\varepsilon^2 \rightarrow 0$, then in this case from Eq. (10), we know that

$$\hat{\beta} = \frac{-S_{XY}S_{YZ} + S_{XZ}S_{YY}}{S_{XX}S_{YY} - S_{XY}^2}. \quad (11)$$

(b) $\lambda \rightarrow \infty$ when $\sigma_\tau^2 \rightarrow 0$, in this case from Eq. (11), we know that

$$\hat{\beta} = \frac{-S_{ZZ}S_{YY} - S_{YZ}^2}{S_{XZ}S_{YY} + S_{XY}S_{YZ}}. \quad (12)$$

(c) When both $\sigma_\varepsilon^2 > 0$ and $\sigma_\tau^2 > 0$, then

$$\hat{\beta} = \frac{k_1 \pm \sqrt{k_1^2 + 4\lambda k_O}}{2\lambda k_O}. \quad (13)$$

When β is determined, γ can also be estimated. After both β and γ are estimated, then α can be estimated by

$$\hat{\alpha} = \bar{Z} - \hat{\beta}\bar{X} - \hat{\gamma}\bar{Y}. \quad (14)$$

(Case ii) When Z , X , and Y are all observed with errors

(a) Both σ_ε^2 and σ_η^2 are known.

We can obtain the two normal equations as follows:

$$S_{XZ} = \beta(S_{XX} - \hat{\sigma}_1^2) + \gamma S_{XY} \quad (15)$$

$$S_{YZ} = \beta S_{XZ} + \gamma(S_{YY} - \hat{\sigma}_\eta^2) \quad (16)$$

Solving Eqs. (15) and (16) by Cramer's rule we have

$$\hat{\beta} = \frac{S_{XZ}S_{YY} - S_{XY}S_{YZ} - S_{XZ}\hat{\sigma}_\eta^2}{S_{XX}S_{YY} - \hat{\sigma}_\varepsilon^2 S_{YY} - \hat{\sigma}_\eta^2 S_{XX} + \hat{\sigma}_\varepsilon^2 \hat{\sigma}_\eta^2 - (S_{XY})^2} \quad (17)$$

$$\hat{\gamma} = \frac{S_{YZ}S_{XX} - S_{XZ}S_{YY} - S_{YZ}\hat{\sigma}_\varepsilon^2}{S_{XX}S_{YY} - \hat{\sigma}_\varepsilon^2 S_{YY} - \hat{\sigma}_\eta^2 S_{XX} + \hat{\sigma}_\varepsilon^2 \hat{\sigma}_\eta^2 - (S_{XY})^2}.$$

(b) Both $\hat{\sigma}_\eta^2 = \lambda_1 \hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\tau^2 = \lambda_2 \hat{\sigma}_\varepsilon^2$ are known.

We can obtain β estimator as one of real roots of the following cubic equation

$$\hat{\beta}^3 H_3 + \hat{\beta}^2 H_2 + \hat{\beta} H_1 + H_O = 0, \quad (18)$$

where

$$\begin{aligned}
 H_3 &= S_{XY}S_{YZ}^2 - MS_{XZ}S_{YZ} - \lambda_1 S_{XY}S_{XZ}^2 \\
 H_2 &= -S_{YZ}^2 + 2\lambda_2 S_{XY}^2 S_{YZ} + MTS_{YZ} - M\lambda_2 S_{XY}S_{YZ} - \lambda_1 S_{XZ}^2 S_{YZ} + 2\lambda_1 TS_{XZ}S_{XY} \\
 H_1 &= \lambda_2^2 S_{XY}^3 - 2\lambda_2 S_{XY}S_{YZ}^2 + MT\lambda_2 S_{XY} + \lambda_2 S_{XY}S_{YZ} + \lambda_2 MS_{YZ}S_{XZ} \\
 &\quad - \lambda_1 \lambda_2 S_{XZ}^2 S_{XY} + 2\lambda_1 S_{XY}S_{XZ}^2 - \lambda_1 T^2 S_{XY} \\
 H_0 &= \lambda_2^2 S_{YZ}S_{XY}^2 + \lambda_2^2 S_{XZ}S_{XY} + \lambda_1 \lambda_2 S_{XZ}^2 S_{YZ} + T\lambda_1 \lambda_2 S_{XZ}S_{YZ} - 2\lambda_1 \lambda_2 TS_{XZ}S_{XY}
 \end{aligned}$$

When $\lambda_2 = 0$, Eq. (18) will reduce to a quadratic equation in $\hat{\beta}$.

100.3.1.3 The Constrained Classical Method

Under the classical case (Case ii), if we only know $\hat{\sigma}_\eta = \hat{\sigma}_\epsilon \lambda$, then we can identify β and γ by imposing $\beta + \gamma = 1$.

We can obtain a quadratic equation in $\hat{\beta}$

$$\hat{\beta}^2(1 - \lambda)S_{XY} + \hat{\beta}(S_{YZ} - \lambda S_{XZ} + 2\lambda S_{XY} + S_{YY} - \lambda S_{XX}) + \lambda(S_{XZ} - S_{XY}) = 0. \quad (19)$$

When $\lambda = 0$, Eq. (19) will reduce to

$$\hat{\beta} = -\frac{S_{YZ} + S_{YY}}{S_{XY}}. \quad (20)$$

Imposing $\beta + \gamma = 1$, upon a multiple regression will help to identify the regression coefficients, but it should also be realized that the constrained regression technique will bias the estimates of the regression coefficients if the unrestricted estimator fails to satisfy the restriction $\beta + \gamma = 1$. The advantages and disadvantages of the constrained regression technique have been discussed by Theil (1971) in some detail.

100.3.2 Grouping Method

Following the structural relationship described in Eq. (1)

$$V_i = \alpha + \beta U_i. \quad (1)$$

Both V_i and U_i are unobserved, and only $Y_i = V_i + \eta_i$ and $X_i = U_i + \epsilon_i$ can be observed. There exists EIV bias when using Y_i and X_i , to investigate the relationship between V_i and U_i .

Wald (1940) proposes a two-portfolio grouping method in dealing with the EIV problem when both dependent and independent variables are subject to measurement errors. He suggests that the measurement error can be reduced by grouping

observations into portfolios. In Wald's two-portfolio grouping method, he groups the independent variable either in descending or ascending order, and divides the observations into two equal groups for both dependent and independent variables; therefore, the first-step estimator of the market model, estimated beta risk, can be written as:

$$\hat{\beta} = \frac{(\bar{Y}_1 - \bar{Y}_2)}{(\bar{X}_1 - \bar{X}_2)}, \quad (21)$$

where $\bar{X}_1$ and $\bar{X}_2$ are the arithmetic means of independent variables for the first and the second groups, respectively; and $\bar{Y}_1$ and $\bar{Y}_2$ are the arithmetic means of dependent variables for the first and the second groups, respectively.

Grouping method is widely used in finance-related research. For example, to minimize the EIV problem in testing the asset pricing model, Black et al. (1972), Blume and Friend (1973), Fama and MacBeth (1973), and Litzenberger and Ramaswamy (1979) use two-pass procedure and the k -portfolio grouping method to examine the capital asset pricing model. By combining securities into portfolios, most of the firm-specific component of the returns can be diversified away and the precision of the beta estimates will be enhanced. The grouping method can, therefore, mitigate the problem raised from measurement errors in estimated beta.

However, some limitations affect the grouping method. First, the grouping method shrinks the range of estimators in the first step and reduces statistical power. To mitigate this problem, in two-pass procedure, the grouping method suggests sorting securities on the first-pass estimator first. Then portfolios are formed by grouping securities with same level of first-pass estimators. This sorting procedure is now standard in empirical tests. Second, a trade-off exists between the bias and the variance of the first-pass estimator according to the number of portfolios. Shanken (1992) argues that the grouping method may cause a larger variation in the portfolio beta. As the number of portfolios (N) increases, the magnitude of the bias becomes greater while the variance of the estimator becomes smaller, and vice versa; therefore, an optimal number of portfolios might exist in which a minimum mean squared error can be obtained. More specifically, when risk premium is estimated by the time-series mean of the cross-sectional regression estimates in testing capital asset pricing model, the mean squared error of the risk premium estimate would be dominated by its bias because its variance would monotonically decrease as the testing period becomes longer. Third, the formation of portfolios for the second-pass estimation might cause a loss of valuable information about cross-sectional behavior among individual securities, because the cross-sectional variations would be smoothed out. Fourth, Ahn et al. (2009) argue that the grouping method, although mitigating measurement error, may yield different results by using different portfolio grouping methods.

Although the grouping method suffers from the limitations discussed above, it still has some clear advantages. With the cross-sectional regression in the second pass, interpreting the results in economic terms is straightforward. Examining model misspecification by testing whether firm characteristics, such as firm size and book-to-market ratio, can explain returns across firms is also convenient. Moreover, the

grouping method is intuitive and easy to implement with real data. The grouping method is therefore still preferred in many empirical studies.

100.3.3 Instrumental Variable Method

Durbin (1954) proposes an instrumental variable method to deal with the EIV problem in a regression model. In the instrumental variable method, the instrumental variable, T_i , is an observable variable known to correlate with V_i and U_i , but is independent of η_i and ε_i . Then β can be estimated by

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^n (T_i - \bar{T})(Y_i - \bar{Y})}{\sum_{i=1}^n (T_i - \bar{T})(X_i - \bar{X})} \\ &= \frac{\sum_{i=1}^n (T_i - \bar{T})(U_i - \bar{U}) + \sum_{i=1}^n (T_i - \bar{T})(\eta_i - \bar{\eta})}{\sum_{i=1}^n (T_i - \bar{T})(V_i - \bar{V}) + \sum_{i=1}^n (T_i - \bar{T})(\varepsilon_i - \bar{\varepsilon})}.\end{aligned}\quad (22)$$

If $\text{plim} \sum_{i=1}^n (T_i - \bar{T})(U_i - \bar{U})$ exists, then $\hat{\beta}$ is a consistent estimator of β because both ε_i and η_i are independent of T_i . Equation (22) can be written in matrix form as follows:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (\mathbf{T}'\mathbf{X})^{-1}\mathbf{T}'\mathbf{Y}, \quad (23)$$

$$\begin{aligned}\text{where, } \mathbf{T}' &= \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ T_1 & T_2 & T_3 & \cdots & T_n \end{bmatrix} \\ \mathbf{X}' &= \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X_n \end{bmatrix}, \text{ and} \\ \mathbf{Y}' &= [Y_1 \ Y_2 \ Y_3 \ \cdots \ Y_n].\end{aligned}$$

However, finding an instrumental variable uncorrelated with η_i and ε_i while highly correlated with V_i and U_i is difficult. Durbin (1954) suggests that if the order of U_i is the same as the order of X_i , then a better instrumental variable would be $T_i = i$, where X_i are ordered by magnitude. That is, $T' = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 3 & \cdots & n \end{bmatrix}$ and $X' = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X_n \end{bmatrix}$. This variable will lead to a more efficient estimate than that of the method of grouping. If we let $T_i = 1$ for X_i greater than its median, $T_i = 0$ for X_i equal to its median, and $T_i = -1$ for X_i smaller than its median, then the

estimator of the instrumental variable method will be the same as the estimator of Wald's two-group grouping method. Therefore, Wald's two-group grouping method is a special case of the instrumental variable method. In other words, the instrumental variable method is more generalized than the grouping method.

Griliches and Hausman (1986) propose an instrumental variable approach to reduce the bias resulted from measurement error. In a panel data framework, they show that instrumental variables estimator is consistent if the measurement error $\varepsilon_{i,t}$ is *i.i.d* across i and t and unobserved independent variable is serially correlated. An instrumental variable using the lags or the difference of lags of the unobserved independent variable can result in a consistent estimator when T is finite and N approaches to infinite.

However, Griliches and Hausman's *i.i.d* assumption is too strong. Biorn (2000) further relaxes Griliches and Hausman's *i.i.d* assumption, and, instead, assumes that the regressor has τ period moving average process. Biorn (2000) shows that using the lags of the variables at least $\tau - 2$ periods as instruments can clear the memory of the moving average process and obtain the consistent estimator.

Lewbel (1997) show simple functions of the model data can be used as instruments for two-staged least squares (TSLS) estimation. Such instruments can be used for identification and estimation when no other instruments are available or improve efficiency.

Given the standard linear regression model with measurement error:

$$Y_i = a + b'W_i + cX_i + e_i, \text{ and} \quad (24)$$

$$Z_i = d + X_i + v_i, \quad (25)$$

in which Y_i , W_i , and Z_i are observable for $i = 1, \dots, n$, while X_i , e_i , and v_i are unobservable. Eqs. (24) and (25) imply that

$$Y_i = \alpha + b'W_i + cZ_i + \varepsilon_i. \quad (26)$$

However, since both Z_i and ε_i depend on v_i , estimators of b and c from OLS regression is inconsistent. Lewbel (1997) shows that the consistent estimators can be obtained by using TSLS with instruments 1 , W_i , and q_i , where q_i is some vector of instruments that are correlated with X_i but not correlated with e_i and v_i .

Lewbel (1997) further empirically applies the instrumental variable method to testing elasticity of patent applications with respect to research and development (R&D) expenditures. He finds, using the TSLS instrumental variable model, the estimated elasticity yields very close to one. Therefore, the TSLS instrumental variable model can mitigate the effects of measurement error and confirm the relationship between patent and R&D.

In addition, Erickson and Whited (2000, 2002) propose a two-step generalized method of moments (GMM) estimators that exploit over-identifying information contained in the high-order moments of residuals obtained from perfectly measured regressors. Basing GMM estimation on residual moments of more than second order requires that the GMM covariance matrix be explicitly adjusted to account for the fact that estimated residuals are used instead of true residuals defined by population

regressions. Erickson and Whited (2000) show that estimators obtained by using moments up to seventh order perform well in Monte Carlo simulations.

Almeida et al. (2010) use Monte Carlo simulations and empirically test investment models to compare the performance of the instrumental variables approach suggested by Biorn (2000) and generalized method of moments. They find that the instrumental variable method can obtain more consistent and efficient estimators than generalized method of moments when independent variables are subject to measurement error.

However, it is difficult to obtain appropriate instrument variables, resulting in weak evidence in empirical research. Lewbel (2012) proposes a new method to deal with measurement error problems in regression model when instrumental variables are not available. Under the assumption of heteroscedastic errors, Lewbel (2012) shows that the regression model with measurement regressors can be identified and estimated by TSLS or GMM.

100.3.4 Mathematical Method

100.3.4.1 Bivariate Case

Deming (1943), York (1966), and Clutton-Brock (1967) have developed a weighted-regression-method-under-iteration approach. Deming (1943) proposed that the best straight line of Eq. (1) can be obtained by minimizing the sum in the following equation:

$$S = \sum_i \left\{ w(X_i) (\hat{U}_i - X_i)^2 + w(Y_i) (\hat{V}_i - Y_i)^2 \right\} \quad (27)$$

$\hat{U}_i$ and $\hat{V}_i$ are the adjusted value of X_i and Y_i , which make the sum in Eq. (27) a minimum. Since we require $\hat{U}_i$ and $\hat{V}_i$ to lie on the best straight line, we must have

$$\hat{V}_i = \alpha + \beta \hat{U}_i, \quad (i = 1, \dots, n) \quad (28)$$

Both $w(X_i)$ and $w(Y_i)$ are the weights of various observations. They are reciprocally proportional to the variance of their measurement error, respectively.

If these values of $\hat{U}_i$, $\hat{V}_i$, α , and β make S a minimum, we have

$$\begin{aligned} & \beta^3 \sum_i \frac{k_i^2 x_i^2}{w(X_i)} - 2\beta^2 \sum_i \frac{k_i^2 x_i y_i}{w(X_i)} - \beta \left\{ \sum_i k_i x_i^2 - \sum_i \frac{k_i^2 y_i^2}{w(X_i)} \right\} + \sum_i k_i x_i y_i \\ & = 0, \end{aligned} \quad (29)$$

where $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$,

$$\bar{X} = \frac{\sum_i k_i X_i}{\sum_i k_i}, \quad \bar{Y} = \frac{\sum_i k_i Y_i}{\sum_i k_i}$$

and

$$k_i = \frac{w(X_i)w(Y_i)}{\beta^2 w(Y_i) + w(X_i)}.$$

Equation (29) is the least-square cubic derived by York (1966). To solve Eq. (29), an initial value is assigned to β to estimate k_i . After obtaining the roots of Eq. (29), one of the legitimate solutions is assigned to estimate k_i and obtain new solutions for β again. A similar procedure is employed iteratively until a convergent solution is obtained.

The mathematical approach involves the estimation of the parameters of a function conditional on the maximum likelihood function adjusted for the true values. This method is different from the classical method in three ways. First, variances of measurement errors for every observation are different. Second, a weighted regression method is applied. Third, the iteration procedure is used to obtain a consistent estimator.

It can be proved that the mathematical programming method reduces to the classical method under three certain conditions.

- (i) Only Y_i has an EIV problem

We can put more weight on X_i which has no EIV problem, $w(X_i) = \infty$, $k_i = w(Y_i)$. We can therefore solve the least square cubic

$$\beta = \frac{\sum_i w(Y_i) x_i y_i}{\sum_i w(Y_i) x_i^2}, \quad (30)$$

which is the estimated coefficient of weighted regression of Y_i on X_i .

- (ii) Only X_i has an EIV problem

In this case, we put more weight on Y_i which has no EIV problem, then $w(Y_i) = \infty$, $k_i = \frac{w(X_i)}{\beta^2}$. We can solve the least square cubic

$$\beta = \frac{\sum_i w(X_i) y_i^2}{\sum_i w(X_i) x_i y_i}, \quad (31)$$

which is the inverse estimated coefficient of weighted regression of Y_i on X_i .

- (iii) Both X_i and Y_i have EIV problem, and $w(X_i)/w(Y_i) = c$.

The least square cubic becomes

$$\beta^2 + \beta \frac{\{c \sum_i k_i x_i^2 - \sum_i k_i y_i^2\}}{\sum_i k_i x_i y_i} - c = 0 \quad (32)$$

100.3.4.2 Multivariate Case

Lee (1973) extends the bivariate mathematical programming method, which was developed by Deming (1943), York (1966), and Clutton-Brock (1967), to a trivariate

case. We define $w(Z_i)$, $w(X_i)$, and $w(Y_i)$ which are the weights of the various observations of Z_i , X_i , and Y_i . It is assumed W , U , and V are functionally rather than structurally related. The mathematical programming procedure begins by minimizing³

$$S = \sum_i \left\{ w(X_i)(x_i - X_i)^2 + w(Y_i)(y_i - Y_i)^2 + w(Z_i)(z_i - Z_i)^2 \right\} \quad (33)$$

s.t. $z_i = \alpha + \beta x_i + \gamma y_i$.

This extension will reduce to Deming's (1943) weighted regression results when the quadratic term of equations are omitted, while Lee's (1973) result is more general than Deming's weighted multiple regression analysis.

100.3.5 Maximum Likelihood Method

In testing capital asset pricing model with dividend and tax, Litzenberger and Ramaswamy (1979) use maximum likelihood method to reduce the effect of errors-in-variables. Litzenberger and Ramaswamy (1979) show that, assuming that the variance of the measurement error in beta is known, the cross-sectional variance of true betas can be replaced by the difference in the variation of the observed betas and the variance of the measurement error. Then the estimator in capital asset pricing model test, under such condition, is consistent by maximum likelihood method.

Kim (1995, 1997, 2010) further provides a maximum likelihood method to correct the EIV problem in testing the asset pricing model. Based upon two-pass capital asset pricing model, Kim (1995) shows that in a multifactor asset pricing model test the EIV leads to an underestimation of the independent variable with a measurement error and an overestimation of the independent variable without measurement error. To correct EIV biases, Kim (1995) extracts additional information about the relation between idiosyncratic error variance, which can be obtained from the first step and the measurement error variance, and incorporates such additional information into the second step of the capital asset pricing model test. Assuming the homoscedasticity of the disturbance term of the market model, Kim (1995) shows that the corrected factors for the traditional least squares estimators of the cross-sectional regression coefficients can be obtained by the maximum likelihood method. The closed form estimators of the multifactor asset pricing model test can therefore be obtained. Assuming the first and second steps of the multifactor asset pricing model are

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + e_{i,t}, \quad \text{and} \quad (34)$$

$$R_{i,t} = \gamma_{0,t} + \gamma_{1,t} \hat{\beta}_{i,t-1} + \gamma_{2,t} V_{i,t-1} + e_{i,t}, \quad (35)$$

³Please see Lee (1973) for the solution of Eq. (33).

where $\beta_{i,t-1}$ is the market risk factor with measurement error, and $V_{i,t-1}$ is a risk factor with no measurement error for security i at time $t-1$. The adjusted estimators in the second step can be written as follows:

$$\begin{aligned}\hat{\gamma}_{1t} &= \frac{M + \left[M^2 + 4\delta_t m_{R\beta}^2 \left(1 - \left(\hat{\rho}_{RV} \hat{\rho}_{\beta V} / \hat{\rho}_{R\beta} \right)^2 \right) \right]^{1/2}}{2m_{R\beta} \left(1 - \left(\hat{\rho}_{RV} \hat{\rho}_{\beta V} / \hat{\rho}_{R\beta} \right) \right)} \\ \hat{\gamma}_{2t} &= \left(m_{RV} - \hat{\gamma}_{1t} m_{\beta V} \right) / m_{VV} \\ \hat{\gamma}_{0t} &= \tilde{R}_t - \hat{\gamma}_{1t} \tilde{\beta}_{t-1} - \hat{\gamma}_{2t} \tilde{V}_{t-1}\end{aligned}\quad (36)$$

where $M = m_{RR} \left(1 - \hat{\rho}_{RV}^2 \right) - \delta_t m_{\beta\beta} \left(1 - \hat{\rho}_{\beta V}^2 \right)$,

$$\begin{aligned}m_{xy} &= (1/N) \sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^N \sum_{j=1}^N w_{ij}, \\ \bar{x} &= \sum_{i=1}^N \sum_{j=1}^N w_{ij} x_i / \sum_{i=1}^N \sum_{j=1}^N w_{ij}, \\ \bar{y} &= \sum_{i=1}^N \sum_{j=1}^N w_{ij} y_i / \sum_{i=1}^N \sum_{j=1}^N w_{ij},\end{aligned}$$

w_{ij} is the (i, j) element of inverse matrix of residual variance in the first step, $\hat{\Sigma}_e^{-1}$, and

$$\hat{\rho}_{xy}^2 = m_{xy} / (m_{xx} m_{yy})^{1/2}.$$

As a result, the maxima likelihood method can correct the problem on exaggerating the estimated coefficient associated to the variable without measurement error. Moreover, the absolute value of estimated intercept by maxima likelihood method is generally smaller than the absolute value of estimated intercept by traditional least squares.

100.3.6 LISREL and MIMIC Methods

Goldberger (1972) conceptually described the LISREL model as a combination of factor analysis and econometrics model. In addition, Anderson (1963) has shown that factor analysis is a generalized version of errors-in-variables (EIV) methods. In this section, we will review and discuss how LISREL and multiple indicators and multiple causes (MIMIC) methods can be used to deal with EIV in finance research.

The linear simultaneous equation system is widely used in finance and accounting-related research. However, a serious limitation of the simultaneous equation approach is an EIV problem. For example, the theoretical determinants of

capital structure in corporate finance can be attributed to unobservable constructs that are usually measured in empirical studies by a variety of observable indicators or proxies. These observable indicators or proxies can then be viewed as measures of latent variables with measurement errors. Maddala and Nimalendran (1996) show that the use of these indicators as theoretical explanatory variables may cause EIV problems. Bentler (1983) also emphasizes the estimated results of the traditional simultaneous equation model has no meaning when variables have measurement errors. Therefore, the latent variable covariance structure model is provided and applied in corporate finance. Titman and Wessels (1988), Chang et al. (2009), and Yang et al. (2009) mitigate the measurement problems of proxy variables, and apply structure equation models (e.g. LISREL model and MIMIC model) to determine capital structure decision. Maddala and Nimalendran (1996) use the structure equation model to examine the effect of earnings surprises on stock prices, trading volumes, and bid-ask spreads.

Goldberger (1972) and Jöreskog and Goldberger (1975) developed a structure equation model with multiple indicators and multiple causes of a single latent variable, MIMIC mode, and obtained maximum likelihood estimates of parameters. Figure 1 shows the path diagram that depicts a simplified MIMIC model in which variables in a rectangular box denote observable variables, while variables in an oval box are latent constructs. In this diagram, observable variables X_1 , X_2 , and X_3 are causes of the latent variable η , while Y_1 , Y_2 , and Y_3 are indicators of η . In our study, X s are determinants of capital structure (η), which are then measured by Y s.

Jöreskog and Sörbom (1989) show that the full structural equation (LIEREL model) can be restricted to a MIMIC model. We here discuss the structural model and show how structural model can be restricted to a MIMIC model.

100.3.6.1 Structural Model (LISEREL Model)

A structural equation model is composed of two sub-models – structural sub-model and measurement sub-model. The structural model can be defined as.

$$\eta = \Gamma X + \zeta, \quad (37)$$

$$Y = \Lambda_y \eta + \varepsilon, \quad (38)$$

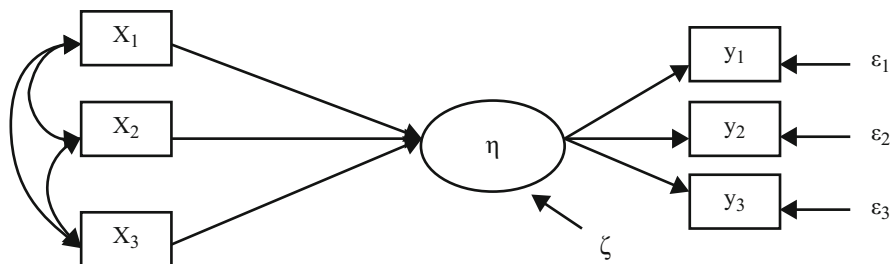


Fig. 1 Path diagram of a simplified MIMIC model

where Y is a vector of indicators of the latent variable η , and X is a vector of causes of η .

The latent variable η is linearly determined by a set of observable exogenous causes, $X = (x_1, x_2, \dots, x_q)'$, and a disturbance ζ . The latent variable η , in turn, linearly determines a set of observable endogenous indicators, $Y = (y_1, y_2, \dots, y_p)'$ and a corresponding set of disturbance, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)'$.⁴

100.3.6.2 MIMIC Model

Substituting Eq. (38) into Eq. (39), we obtain a reduced form:

$$Y = \Lambda_y \eta + \varepsilon = \Lambda_y (\gamma' X + \zeta) + \varepsilon = \Pi' X + z \quad (39)$$

In structural equation modeling, the total effect of a cause variable on an indicator can be measured as the sum of the direct effect and the indirect effect. Since a MIMIC model is a reduced form of a structural equation model, the total effect of MIMIC model, denoted as Π' in Eq. (39), comes merely from the indirect effect.

Since the scale of the latent variable is unknown, the factor indeterminacy is a common problem in the MIMIC model, as in other structure equation models. We can obtain infinite parameter estimates from the reduced form by arbitrarily changing the scale of the latent variables. However, by fixing the scales of latent variables, one can solve the indeterminacy problem. Two methods are usually adopted to fix the scale of latent variables. One method is normalization in which a unit variance is assigned to each latent variable, while another method is to fix a nonzero coefficient at unity for each latent variable.

In terms of estimation of the parameters, Jöreskog and Goldberger (1975) adopt the normalization method to deal with the factor indeterminacy problem and use maximum likelihood estimation method in structural equation modeling to estimate parameters. The maximum likelihood estimates for the parameters of the model are obtained at the minimization of the fit function as follows:

$$F = \log \|\Sigma\| + \text{tr}(S\Sigma^{-1}) - \log \|S\| - (p-q), \quad (40)$$

where Σ is the population covariance matrix; S is the model-implied covariance matrix; p is the number of exogenous observable variables; and q is the number of endogenous observable variables. Minimization of the fit function can be done by the LISREL program provided by Jöreskog and Sörbom (1981).

100.3.7 Bayesian Approach

Zellner (1970) uses the Bayesian approach to deal with measurement problems in the estimation of regression relationships containing unobservable independent

⁴Stapleton (1978) further develops MIMIC with more latent variables.

variables. Zellner (1970) shows that the Bayesian approach can obtain optimal estimates under a finite sample. Several studies use Bayesian approaches to examine cost of capital (e.g., Lee and Wu 1989) and asset pricing models (e.g., Ang and Chen 2007; Davis 2010; Geweke and Zhou 1996; McCulloch and Rossi 1991).

Davis (2010) develops a Bayesian approach and uses US firm level data to reexamine the capital asset pricing model. The Bayesian approach can estimate all parameters simultaneously in one step and effectively avoid the errors-in-variables problem on the estimators induced from two-pass capital asset pricing test.

Davis (2010) uses a Bayesian approach to simultaneously estimate coefficients of the following three equation system.

$$r_{i,t,y} = \alpha_{i,y} + \gamma_{i,y}r_{m,t,y} + \delta_{i,y}r_{m,t-1,y} + \varepsilon_{i,t,y}, \text{ where } \varepsilon_{i,t,y} \sim N(0, \sigma_{r_{i,y}}^2); \quad (41)$$

$$\overline{r_{i,y}} - \overline{r_{f,y}} = c_{0,y} + c_{m,y}\beta_{i,y} + \eta_{i,y}, \quad (42)$$

where $\eta_{i,y} \sim N(0, \sigma_y^2)$ and $\beta_{i,y} \sim \gamma_{i,y} + \delta_{i,y}$;

$$\mathbf{c}_y = \begin{bmatrix} c_{0,y} \\ c_{m,y} \end{bmatrix} \tilde{N}(\mathbf{c}, \mathbf{V}_c), \text{ where } \mathbf{c} = \begin{bmatrix} c_0 \\ c_m \end{bmatrix} \text{ and } \mathbf{V}_c = \begin{bmatrix} \sigma_0^2 & \sigma_{0m}^2 \\ \sigma_{0m}^2 & \sigma_m^2 \end{bmatrix}; \quad (43)$$

where $r_{i,t,y}$ is firm i 's return in month t during the time period y , and $\overline{r_{i,y}} - \overline{r_{f,y}}$ denotes the average monthly excess return for firm i during the time period y . The model allows firm-level β s to vary over each time period, y . The model also assumes that the joint normality of stock returns and market returns, contemporaneously estimated β s, and average excess returns are statistically independent.

In addition to dealing with errors-in-variables problem, there are several advantages using the Bayesian approach to test the capital asset pricing model. First, the Bayesian approach allows β s to vary over time periods and firms and controls the inherent uncertainty associated with firm-level β s. Second, the Bayesian approach can modify the distribution assumptions in stock returns and market returns. Third, the Bayesian inference is free from the use of asymptotic approximations and therefore can be used under finite sample. Fourth, the Bayesian approach takes parameter uncertainty associated with all the model parameters into account.

100.4 Applications of Errors-in-Variables Models in Finance Research

For the last four and a half decades, EIV models have been used to correct estimation bias associated with empirical results in various finance-related research issues. We here review four kinds of research, cost of capital, asset pricing models, capital structure, and investment equation, and discuss how EIV models can remedy measurement error problems induced from finance-related research. It is therefore

useful to understand the statistical properties of these EIV models in situations resembling real research question.

The main focus of this section is to discuss the measurement error problems on various empirical studies related to finance research and investigate how EIV models can remedy such problems. However, in empirical studies, it is impossible to observe variables without measurement error. We cannot evaluate EIV models and suggest a best EIV model for a certain circumstance. Instead, we here provide Table 1 to summarize the application of EIV models in finance-related research. Research topics, EIV models, specialties of EIV model, and results for each study are included in Table 1.

100.4.1 Cost of Capital

Miller and Modigliani (1966) developed a theoretical expression for the value of a firm from which the firm's cost of capital could be derived. They assume a perpetual stream of earnings from real assets, and a constant capitalization rate (ρ), at which the market discounts the uncertain pure (unlevered) equity stream of earnings for some risk classes and perfect markets. It is thus possible to estimate the market capitalization rate (and thus the cost of capital) of a group of firms by performing a cross-sectional regression of the market value of the firm's equity on the expected average earnings of the firm, the market value of debt, and the growth resulting from the above-average investment opportunities. The above analysis suggests a cross-sectional regression:

$$(V - \tau_c D) = a_0 + a_1 \bar{X}(1 - \tau_c) + a_2(\text{growth potential}) + \varepsilon, \quad (44)$$

where V is sum of the market value of all securities issued by the firm, τ_c is the corporate tax rate, D is the market value of a firm's debt, and $\bar{X}$ is the expected level of average annual earnings generated by current assets.

To avoid heteroscedasticity of regression residuals, the equation must be adjusted to compensate for the dominance of the large companies. Miller and Modigliani (1966) use weighted least square to adjust the standard deviation of the error term to firm size (deflating each variable by the book value of total assets). Therefore Eq. (44) can be adjusted to:

$$\frac{(V - \tau_c D)}{A} = \frac{a_0}{A} + a_1 \bar{X} \frac{(1 - \tau_c)}{A} + a_2 \frac{\Delta \bar{A}}{A} + u, \quad (45)$$

where $u = \varepsilon / A$. With this reformulation, the regression equation is expected to be homogeneous, that is, to have no constant term, and the term A , total assets, is used to avoid heteroscedasticity.

An additional problem beyond that of heteroscedasticity is the possible error of measurement associated with the earnings term. Since anticipated average earnings are essentially unobservable, accounting-statement estimates of earnings must be

Table 1 Applications of errors-in-variables models in finance research

Study	Issue	Method	Specialties/ conditions	Results
Miller and Modigliani (1966)	Determinants of cost of capital	Instrumental variable method	Have to find exogenous variables	Measurement error problem matters
Black et al. (1972)	CAPM test	Grouping (10 groups)	Panel data; 2-pass estimation	Reject both the CAPM and the zero-beta CAPM
Blume and Friend (1973)	CAPM test	Grouping (12 groups)	Panel data; 2-pass estimation	Linear model is better than quadratic model in explaining expected return Reject both the CAPM and the zero-beta CAPM
Fama and MacBeth (1973)	CAPM test	Grouping (20 groups), period by period	Panel data; 2-pass estimation	Find a linear relationship between the expected return and beta risk, beta is the only risk measure in explaining expected return, and risk premium is greater than zero CAPM and efficient capital market hold
Higgins (1974)	Determinants of cost of capital	Smoothing procedure	Assume that measurement error has zero mean and serially uncorrelated	The extrapolation of historical population trends is superior to the conventional use of change of capital, and share prices are not a positive function of dividends as often suggested
Lee (1977)	CAPM test	Wald's grouping/ instrumental variable	Adjust for measurement error of market return in first-step	Estimated risk premium is larger than realized risk premium Reject CAPM
Litzenberger and Ramaswamy (1979)	CAPM test	MLE, OLS, GLS	For individual stocks	Before-tax expected rates of return are linearly related to systematic risk and dividend yield MLE can obtain consistent estimators without losing

(continued)

Table 1 (continued)

Study	Issue	Method	Specialties/ conditions	Results
				efficiency CAPM is rejected because of nonzero $\hat{\gamma}_0$
Cheng and Grauer (1980)	CAPM test	Grouping (20 groups)	Price-level testing (invariance law)	Neither framework of invariance law nor security market line can accommodate the possibility that the CAPM may hold for each period Reject CAPM
Gibbons (1982)	CAPM test	One-step Gauss- Norman procedure (40 groups)	One-step estimation	Gauss-Norman procedure can increase the precision of estimated risk premium Reject CAPM
Titman and Wessels (1988)	Determinants of capital structure	LISREL model	Deal with EIV problem due to the imperfect representation of proxy variables for interested attributes	Do not support for four of eight propositions on the determinants of capital structure A firm's capital structure is not significantly related to its non-debt tax shields, volatility of earnings, collateral value of assets, and future growth
Lee and Wu (1989)	Determinants of cost of capital	Zellner's EV method	Use sample information to estimate the ratio of error variances and construct an operational estimator	Obtain better cost of capital estimates
MacKinlay and Richardson (1991)	CAPM test	GMM	Release the assumption of the normality of asset returns	Conclusions of mean-variance efficiency vary by settings
Shanken (1992)	CAPM test	MLE	For individual stocks; deal with small-sample bias in the second-step	The adjustment does not have much effect on Fama and MacBeth's (1973)

(continued)

Table 1 (continued)

Study	Issue	Method	Specialties/ conditions	Results
			cross-sectional regression estimates	conclusion Support CAPM
Fama and French (1992)	CAPM test	2-way grouping (10 × 10 groups)	Take size and book-to-market ratio into account	The market capitalization and the book-to-market ratio can replace beta altogether Reject CAPM
Jagannathan and Wang (1996)	CAPM test	Multifactor Asset Pricing Model	Test conditional CAPM	Including human capital and business cycle can increase explanatory power of expected return Support CAPM
Kim (1995, 2010)	CAPM test	MLE	Closed form adjustment; for individual stocks or groups	MLE method can effectively adjust the errors-in-variables bias and CAPM holds Support CAPM
Kim (1997)	CAPM test	MLE	Closed form adjustment; for individual stocks or groups; for multifactor estimation	Linear relationship between beta and expected return Book-to-market ratio has significant explanatory power for expected return, but size has not
Lewbel (1997)	Elasticity of patent applications to R&D expenses	Instrumental variable method	Applicable if no outside data available for use as instruments	TSLS estimator can mitigate the effects of measurement error The estimated elasticity of patent applications with respect to R&D expenditures yields very close to one
Erickson and Whited (2000)	Test q theory	GMM	For balanced panel data	Cash flow does not affect firms' financial decision, even for financially constrained firms Support the q theory if measurement error is taken into account
Pastor et al. (2008)		Implied cost of capital	Use earnings forecasts to compute	The implied cost of capital can capture

(continued)

Table 1 (continued)

Study	Issue	Method	Specialties/ conditions	Results
	Determinants of cost of capital		implied cost of capital	time variation in expected stock returns
Chang et al. (2009)	Determinants of capital structure	MIMIC model	Allow several observable variables as indicators without multicollinearity problem	Seven constructs, growth, profitability, collateral value, volatility, non-debt tax shields, uniqueness, and industry, as determinants of capital structure have significant effects on capital structure decision
Yang et al. (2009)	Determinants of capital structure	LISREL model	Jointly determine capital structure and return	Stock returns, expected growth, uniqueness, asset structure, profitability, and industry classification are main determinants of capital structure Leverage, expected growth, profitability, firm value, and liquidity can explain stock returns The capital structure and stock return, in addition, are mutually determined by each other
Almeida et al. (2010)	Test q theory	GMM and instrumental variables method	Simple instrumental variable – lagged variable	Estimators from GMM are unstable across different specifications and not economically meaningful Estimators from a simple instrumental method are robust and conform to q theory
Davis (2010)	CAPM test	Bayesian approach	One-step estimation	Positive relationship between excess return and market

(continued)

Table 1 (continued)

Study	Issue	Method	Specialties/ conditions	Results
				risk Support CAPM
Jagannathan et al. (2010)	CAPM tests	Three-stage cross- sectional regression	Adjust for rolling window betas	Dealing with measurement error in rolling window betas Support CAPM
Guay et al. (2011)	Determinants of cost of capital	Corrected implied cost of capital	Sluggish analyst forecasts may result in measurement error on implied cost of capital	The corrected implied cost of capital can improve the ability to explain cross-sectional variation in future stock returns
Da et al. (2012)	CAPM tests	Three-stage cross- sectional regression	Adjust for rolling window betas	Dealing with measurement error in rolling window betas Support CAPM
Erickson and Whited (2012)	Test q theory	GMM	For unbalanced panel data	Instrumental variables, dynamic panel estimators, and high-order moment estimators can perform well under correct specification Developing a minimum distance technique allowing high-order moment estimators be used in unbalanced panel data
Ortiz-Molina and Phillips (2014)	Determinants of cost of capital	Implied cost of capital	Use earnings forecasts to compute implied cost of capital	The implied cost of capital can mitigate measurement error problem on determining the cost of capital
Chen et al. (2019)	Determinants of capital structure	SEM with CFA approach	Jointly determine capital structure and return	SEM with CFA approach outperforms MIMIC model and 2SLS method in terms of the joint determinants of capital structure and stock return

used instead. Therefore, the true relation between value and anticipated earnings, when replaced by the observable estimates, implies a simultaneous system of relationships:

$$V_i^* = \alpha X_i^* + \sum_j \beta_j Z_{ij} + u_i, \quad (46)$$

$$X_i = X_i^* + v_i, \quad (47)$$

$$X_i^* = \sum_j \delta_j Z_{ij} + w_i, \quad (48)$$

where $V_i^* = \frac{V_i - \tau_c D_i}{A_i}$, $X_i^* = \frac{\bar{X}(1 - \tau_c)}{A_i}$ (the true anticipated earnings); v_i = measurement errors associated with current earnings; X_i = observable estimate of earnings derived from the accounting statements; and Z_{ij} = other relevant variables determining earnings. Equations (46), (47), and (48) are related to anticipated earnings and a set of explanatory variables, which may also be correlated with the firm's anticipated earnings.

In addition, the earnings variable used in the regression only approximates the true value of anticipated earnings, varying by the error of measurement, v_i . The system represents the simultaneous determination of two endogenous variables, V^* and X , by the Z_j exogenous variables. In regressing:

$$V_i^* = \alpha X_i + \sum_j \beta_j Z_{ij} + U' \quad (49)$$

the coefficients will be biased. The coefficient for earnings, α , will have a downward bias.

In an attempt to remedy the simultaneous-equation bias, Miller and Modigliani (1966) use an instrumental-variable approach. In this approach, the endogenous variable X is first regressed against all the instrumental variables, Z_j , to obtain estimates of the various coefficients. These estimates are then used to develop a new variable, $\hat{X}$, which is

$$\hat{X}_i^* = \sum_j \hat{\delta}_j Z_{ij} \quad (50)$$

Depending on the choice of Z_j , the new estimate of earnings, $\hat{X}_i^*$, should be relatively free of the error measurement. It can then be used in the second-stage regression as the earnings variable. The resulting estimates of α and β can be shown to be consistent.

Miller and Modigliani (1966) hypothesized that the constant term was really zero. The reduction of bias on the estimates through the use of the two-stage process also seems to support the hypothesis that the constant term is zero. Miller and Modigliani (1966) state that the reason the constant term was significantly different from zero for the direct least-squares cases was that the error of measurement for earnings was large. This error is reduced by the two-stage process.

Higgins (1974) derives and tests a finite-growth model for the estimation of the cost of capital and share price of electric utility industry between 1960 and 1968. He suggests that the market value of equity is related to the trend of earnings and the trend of population in utility's service area. Assuming that observations of a variable consist of a true component and a random element, if such random elements have zero mean and are serially uncorrelated, the smoothing procedure can reduce potential errors in measurement. Empirical results show that the extrapolation of historical population trends is superior to the conventional use of change of capital, and share prices are not a positive function of dividends as often suggested.

Zellner (1970) proposes a least squares regression method to deal with potential errors-in-variables problems. He shows that his methods utilize more information than traditional instrumental variables methods do in dealing with an errors-in-variables problem. Lee and Wu (1989) further apply Zellner's method to reexamine Miller and Modigliani's (1966) cost of capital estimation for utility industry and obtain better cost of capital estimates than OLS methods and instrumental variable method.

More recently, Pastor et al. (2008) propose an implied cost of capital, which is calculated by earnings forecasts, and argue that the implied cost of capital can capture time variation in expected stock returns. Ortiz-Molina and Phillips (2014) adopt Pastor et al. (2008) method to investigate the relationship between real asset liquidity and the cost of capital, and find the implied cost of capital can mitigate measurement error problem on determining the cost of capital. Guay et al. (2011) further propose the implied cost of capital corrected sluggish analyst forecast to improve the ability to explain cross-sectional variation in future stock returns.

100.4.2 Capital Asset Pricing Model

The capital asset pricing model (CAPM) developed by Sharpe (1964), Lintner (1965), and Mossin (1966) implies that the expected returns on securities and their market risks (β) are positively and linearly correlated and that market risks have sufficient power to explain expected returns of securities. Black et al. (1972), Fama and MacBeth's (1973), and others use the two-step method to test CAPM. In the first step, estimated betas are obtained by time-series market model for each security. In the second step, the estimated betas are used in testing the linear relationship between betas and expected returns on securities. Because estimated betas are subjected to a measurement error (estimation error) problem, there exists an EIV problem in the second step. The EIV problem will result in estimating the explanatory power of beta and the estimated rate of return on beta risk. More specifically, the EIV problem leads to an underestimation of the coefficient associated with beta risk. Although the EIV problem exists in the two-step method of the asset pricing test, most researchers do not carefully use sophisticated econometric and statistical methods to deal with this kind of problem. Roll (1969, 1977) shows that the testing asset pricing model suffers an EIV problem, concluding that (i) no correct and unambiguous test of the theory has appeared in the literature, and (ii) practically

no possibility exists that such a test can be accomplished in the future. Roll and Ross (1994) show that the measurement error problem of market rate of return can bias the empirical test of CAPM.

Several studies focus on beta estimation in the first step to solve the EIV problem in testing CAPM. Brennan (1970), Lee and Jen (1978), and Roll (1977) show that the possible measurement error in market beta risk is the unobserved market rate of return and risk-free rate of return.⁵ To improve the beta estimator, Fabozzi and Francis (1978) and Lee and Chen (1979) use the random coefficient procedure to estimate random coefficient betas. Brennan and Schwartz (1977), Brennan (1979), and Brown and Warner (1980) also provide different types of market models that can produce different results in predicting rates of return, testing efficient market hypotheses, and measuring security price performance.

To reduce the impact of the measurement error problem in the second step, Black et al. (1972) use the grouping method in the capital asset pricing model test. Although the results support the linear relationship between the systematic risk and expected return, CAPM cannot hold because of the nonzero intercept term and the lower market premium in the cross-sectional regression. Blume and Friend (1973) and Fama and MacBeth (1973) also use the grouping method in testing CAPM and show that the CAPM is valid. However, Jagannathan et al. (2009) show that the time series average of the cross-sectional estimators converges in probability to the true value of the estimator. Although their results support CAPM indicating a linear relationship between the systematic risk and the expected return, the lower value of time-series average of the market premium shows that the measurement error problem still exists after Fama and MacBeth's (1973) grouping method.

Considering the measurement errors of the market rate of return and risk-free rate of return, Lee (1977) uses two EIV estimation methods, Wald's two-group grouping method and Durbin's instrumental variable method, to adjust the estimated beta risk in the first step of the capital asset pricing test. Although correcting the measurement errors induced from the unobservable market rate of return, Lee (1977) finds that the predictive ability of the capital asset pricing model is still poor.

Litzenberger and Ramaswamy (1979) derive an after-tax version of CAPM and show that, in the equilibrium, the before-tax expected return on a security is linearly related to its systematic risk and its dividend yield. Litzenberger and Ramaswamy (1979) further empirically test both the before tax and the after-tax versions of

⁵Roll (1969, 1977), and Lee and Jen (1978) show that the observed market rate returns in terms of stock market index are measured with errors since the stock market index does not include all assets which investors can invest. Lee and Jen (1978) have theoretically shown how beta estimate and Jensen performance measures can be affected by both constant and random measurement errors of R_m and R_f .

Diacogiannis and Feldman (2011), Green (1986), Roll and Ross (1994), and Gibbons and Ferson (1985) have argued that market portfolio measure with errors is an inefficient portfolio and show how the inefficient benchmark can affect theoretical CAPM derivation. Diacogiannis and Feldman (2011) provide a pricing model that uses inefficient benchmarks, a two beta model, one induced by the benchmark, and one adjusting for its inefficiency.

CAPM. Instead of grouping method, Litzenberger and Ramaswamy (1979) use maximum likelihood estimation in the second-step regression to test the before-tax and the after-tax versions of capital asset pricing model. Although maximum likelihood estimators are consistent, the average risk premium is small and not significantly different from zero.

Given the EIV bias in the two-step CAPM test, Gibbons (1982) introduces a one-step Gauss-Newton procedure and uses the maximum likelihood method to obtain the estimated price of systematic risk. Because the one-step Gauss-Newton procedure does not use estimated beta as an explanatory variable in a regression model, the measurement errors problem of estimated beta can be avoided. Shanken (1992) shows that using generalized least square (GLS) on the second step of CAPM test can yield an estimator identical to the Gauss-Newton estimator obtained by Gibbons' (1982) maximum likelihood method. Gibbons (1982) shows that the Gauss-Newton procedure increases the precision of estimated risk premium, but rejects the mean-variance efficiency of the market portfolio.

Shanken (1992), who provides a modified version of the two-step estimator by using maximum likelihood estimation, finds that Fama and MacBeth's two-step procedure overstates the precision of the estimator in the second-step and therefore provides an adjusted standard error for the estimator in the second-step regression. Jagannathan and Wang (1998a, b) further release Shanken's assumption that asset returns are conditional homoscedasticity to derive a more general standard error for the second-step estimator by generalized least squares.

Fama and French (1992) use a two-way sort grouping method to control for size effect, and find a weak relationship between beta risk and expected return. Before reaching the conclusion that the capital asset pricing has not been valid in the recent years, one possible reason that may be considered is that the measurement error problem cannot be fully eliminated by the grouping method, and results of CAPM test may vary depending on the portfolio formation technique (e.g., Ahn et al. 2009).

To deal with the problem of EIV in testing CAPM, Kim (1995) provides a maximum likelihood method, extracting information associated with the relationship between the measurement error variance and idiosyncratic error variance and incorporating such information into the maximum likelihood estimation in the second step of the capital asset pricing model test. Given the assumption that the disturbance term of the market model is homoscedasticity, the corrected factors for the traditional least squares estimators of the cross-sectional regression coefficients can be obtained. Although Kim's (1995) maximum likelihood method can only deal with the EIV problem of the estimated beta in the first pass, the maximum likelihood method can test, besides the capital asset pricing model, the multifactor asset pricing models. Kim (1995) uses maxima likelihood method to reexamine CAPM and multifactor asset pricing model and finds more support for the role of market beta risk and less support for the role of firm size. His results show that the prominent risk factors (e.g., size, book-to-market ratio, and momentum factors) might result in a different explaining power for cross-sectional stock returns after correcting the EIV problem.

MacKinlay and Richardson (1991) use generalized method of moments (GMM) to test the mean-variance efficiency. They theoretically show that the estimator from GMM and the estimator from maximum likelihood method are equivalent when stock returns are conditionally homoscedasticity, but GMM can avoid the EIV problem by estimating coefficients in one step. Empirical and simulation results show that the conclusion mean-variance efficiency of market indexes is sensitive to the model settings.

Lee et al. (2019) show that, when applying the two-step regression method to test the asset pricing model, the errors-in-variables problem underestimates the market beta and overestimates other risk factors with no measurement error. They, therefore, provide three alternative errors-in-variable models in testing the capital asset pricing model, including the grouping method, the instrumental variable method, and the maximum likelihood method.⁶ Using the grouping method in both first step and second step to deal with measurement error of market return and estimated beta, Lee et al. (2019) show that only the two-group grouping method used in the second step can improve the price of the beta risk, which is market premium (see Table 2).

Table 3 presents results of the instrumental variable method for testing the capital asset pricing model. Uncorrected estimators, estimators corrected in the first step, estimators corrected in the second step, and estimators corrected in both first step and second step are presented in order. Results show that, after correcting by instrumental variables method in the second step, the corrected price of beta risk has been improved both in magnitude and significant level.

Table 4 presents the time series average of the cross-sectional regression estimators by the maximum likelihood method and by traditional ordinary least square method. After correcting the errors-in-variables problem by the maximum likelihood method, for the capital asset pricing model test, the estimated intercept decreases in both magnitude and significant level, and the estimated price of beta risk increases and turns to significant. For the multifactor asset pricing model, the estimated intercept and the estimated coefficient of the size factor decrease in both magnitude and significance level; and the estimated price of beta risk turns to significant. The results therefore show that the evidence of failure in the capital asset pricing model or the finding of new risk factors might occur because of the errors-in-variables problem rather than model misspecification.

Lee et al. (2019) shows that both potential measurement error problems of market return in the first pass and estimated beta in the second pass are corrected by either the grouping method or the instrumental variable method. In addition, the errors-in-variable model can improve the capital asset pricing model tests at the individual stock level. The empirical results support the role of market beta in the capital asset pricing model after correcting the EIV problem.

To deal with the measurement error problem associated with testing both CAPM and APT, Lee and Wei (1984) and Wei (1984) use the MIMC model to test whether

⁶Please check Chap. 8 of Lee et al. (2019) entitled “Three alternative methods in testing capital asset pricing model.”

Table 2 Grouping method in both first step and second step for testing capital asset pricing model This table provides the corrected estimators of the capital asset pricing test by using Wald’s grouping method in both first step and second step. In the first step, 60 time-series observations of each stock are divided into k groups based on their corresponding market returns ($k = 2, 10, 15, 20$, and 30). The market-adjusted preranking beta for each stock is estimated by the market model using k time-series observations. In the second step, based on each firm’s preranking beta, portfolios are formed into N ($N = 2, 10, 20, 50, 100$, and 200) beta-sorted portfolios in month t . Postranking beta of each beta-sorted portfolio in month t is the average of preranking betas of each portfolio. The monthly cross-sectional regression coefficients for each portfolio, $\hat{\gamma}_{0t}$ and $\hat{\gamma}_{1t}$, are estimated through the ordinary least square. Parameters, $\bar{\gamma}_0$ and $\bar{\gamma}_1$, are the time-series averages of the month-by-month cross-sectional regression coefficient estimates by using average beta as postranking portfolio beta. Coefficients are presented in percentages, and t -statistics are presented in the parentheses. The sample period is from January 1931 to December 2009

Portfolio Size (first step/second step)		$R_{it} = \gamma_{0t} + \gamma_{1t}\hat{\beta}_{it-1} + e_{it}$				
		$\bar{\gamma}_0$		$\bar{\gamma}_1$		$\bar{R}^2$
60	N	0.7715	(5.74)	0.3513	(1.90)	0.0329
2	2	0.7544	(6.28)	0.3661	(1.95)	–
2	10	0.8046	(6.41)	0.3169	(1.75)	0.5455
2	20	0.8096	(6.40)	0.3123	(1.73)	0.4373
2	50	0.8197	(6.48)	0.3028	(1.68)	0.2981
2	100	0.8177	(6.45)	0.3056	(1.70)	0.2076
2	200	0.8215	(6.47)	0.3054	(1.71)	0.1339
10	2	0.6881	(5.61)	0.4326	(2.15)	–
10	10	0.7505	(5.97)	0.3713	(1.90)	0.5664
10	20	0.7602	(6.00)	0.3618	(1.87)	0.4546
10	50	0.7675	(6.03)	0.3554	(1.85)	0.3164
10	100	0.7639	(6.00)	0.3585	(1.87)	0.2222
10	200	0.7663	(5.99)	0.3584	(1.86)	0.1453
15	2	0.7173	(5.78)	0.4046	(2.04)	–
15	10	0.7560	(5.91)	0.3668	(1.90)	0.5689
15	20	0.7648	(5.94)	0.3583	(1.87)	0.4573
15	50	0.7660	(5.93)	0.3570	(1.88)	0.3175
15	100	0.7711	(5.98)	0.3511	(1.85)	0.2228
15	200	0.7642	(5.89)	0.3567	(1.89)	0.1454
20	2	0.6857	(5.56)	0.4355	(2.16)	–
20	10	0.7538	(5.85)	0.3688	(1.92)	0.5666
20	20	0.7597	(5.85)	0.3627	(1.90)	0.4551
20	50	0.7630	(5.86)	0.3602	(1.90)	0.3139
20	100	0.7671	(5.89)	0.3542	(1.88)	0.2204
20	200	0.7656	(5.85)	0.3529	(1.88)	0.1438
30	2	0.6886	(5.49)	0.4320	(2.16)	–
30	10	0.7591	(5.83)	0.3631	(1.90)	0.5670
30	20	0.7625	(5.82)	0.3598	(1.90)	0.4544
30	50	0.7654	(5.84)	0.3575	(1.90)	0.3143
30	100	0.7682	(5.83)	0.3550	(1.89)	0.2208
30	200	0.7624	(5.80)	0.3615	(1.92)	0.1439

Table 3 Instrumental variable method for testing capital asset pricing model This table provides corrected estimators of the capital asset pricing test by using Durbin's instrumental variable approach in both first step and second step. In the first step, the instrumental variables for 60 time-series observations of each stock are ranking orders of their corresponding market returns. Then the beta of each security can be estimated by the instrumental variable method. In the second step, the instrumental variables for cross-sectional observations in month t are the rank orders of the estimated betas of cross-sectional observations. The monthly cross-sectional regression coefficients in month t , $\hat{\gamma}_{0t}$ and $\hat{\gamma}_{1t}$, are estimated through the instrumental variable method. Parameters, $\bar{\gamma}_0$ and $\bar{\gamma}_1$, are the time-series averages of the month-by-month cross-sectional regression coefficient estimates. The t -statistics are provided under the null hypothesis that the estimator is different from zero. This table also presents the number of observations and the percentage of observations that are significant under the null hypothesis that the absolute value of corrected estimator is greater than the absolute value of the uncorrected estimator at 95% and 99% significance levels. Coefficients are presented in percentages, and t -statistics are presented in the parentheses. The sample period is from January 1931 to December 2009. Results of subperiods are also provided

Time Period	Method (first step/ second step)	$R_{it} = \gamma_{0t} + \gamma_{1t}\hat{\beta}_{it-1} + e_{it}$								$\overline{R^2}$
		$\widehat{\gamma}_0$	t -stats	Compare to OLS/OLS (95%/99%)	Compare to IV/OLS (95% / 99%)	$\widehat{\gamma}_1$	t -stats	Compare to OLS/OLS (95%/99%)	Compare to IV/OLS (95%/99%)	
1931– 2009	OLS OLS	0.7715	(5.74)				0.3513 (1.90)			0.0329
	IV OLS	0.7884	(6.25)				0.3338 (1.75)			0.0331
	OLS IV	0.7346	(5.77)	18.88%	10.44%		0.3877 (2.02)	21.84%	14.35%	–
	IV IV	0.7491	(6.21)	26.37%	18.04%	11.08%	0.3725 (1.90)	27.95%	20.89%	–
1931– 1950	OLS OLS	0.7655	(1.96)				1.0590 (1.82)			0.0587
	IV OLS	0.6992	(2.02)				1.1224 (1.81)			0.0601
	OLS IV	0.7022	(1.93)	17.08%	10.83%		1.1208 (1.86)	18.33%	12.50%	–
	IV IV	0.6425	(1.96)	28.75%	23.75%	8.75%	1.1774 (1.85)	29.58%	25.42%	–
1951– 1970	OLS OLS	0.8329	(5.09)				0.0466 (0.21)			0.0316
	IV OLS	0.8709	(5.24)				0.0107 (0.05)			0.0310
	OLS IV	0.8200	(5.16)	14.17%	4.17%		0.0605 (0.27)	15.83%	6.25%	–

(continued)

Table 3 (continued)

Time Period	Method (first step/ second step)	$R_{it} = \gamma_{0t} + \gamma_{1t}\hat{\beta}_{it-1} + e_{it}$									
		$\widehat{\gamma}_0$	t -stats	Compare to OLS/OLS (95%/99%)	Compare to IV/OLS (95% / 99%)	$\widehat{\gamma}_1$	t -stats	Compare to OLS/OLS (95%/99%)	Compare to IV/OLS (95%/99%)	$\overline{R}^2$	
1971– 1990	IV	0.8505	(5.29)	19.58%	9.58%	0.0323	(0.15)	20.83%	15.83%	10.42%	
	OLS	0.7575	(3.27)			−0.1441	(−0.52)			0.0192	
	IV	0.7383	(3.24)			−0.1259	(−0.45)			0.0192	
	OLS	0.7292	(3.23)	19.17%	12.92%	−0.1177	(−0.41)	22.08%	15.83%	−	
	IV	0.7035	(3.16)	30.83%	21.67%	−0.0934	(−0.32)	32.92%	23.33%	15.42%	
	OLS	0.7279	(1.07)			0.4486	(0.56)			0.0213	
	IV	0.8482	(1.27)			0.3276	(0.45)			0.0217	
	OLS	0.6846	(1.08)	25.44%	14.04%	0.4925	(0.59)	31.58%	23.25%	−	
	IV	0.8026	(1.27)	26.32%	17.11%	0.3738	(0.49)	28.51%	31.58%	21.49%	

Table 4 Maximum likelihood method for testing capital asset pricing model This table provides the corrected estimators of the capital asset pricing test by using maximum likelihood method. The corrected estimators are the monthly cross-sectional regression coefficients through the maximum likelihood method, and the uncorrected estimators are those estimated through the ordinary least square. Parameters $\tilde{\gamma}_0$, $\tilde{\gamma}_1$, and $\tilde{\gamma}_2$ are the time-series averages of the month-by-month cross-sectional regression coefficients. Coefficients are presented in percentages, and *t*-statistics are presented in parentheses. The sample period runs from January 1931 to December 2009. Results of subperiods are also provided

Time period	$R_{it} = \gamma_{0,t} + \gamma_{1,t}\hat{\beta}_{1,t-1} + e_{i,t}, R_{it} = \gamma_{0,t} + \gamma_{1,t}\hat{\beta}_{1,t-1} + \gamma_{2,t}\log(V_{i,t-1}) + e_{i,t}$					
	MLE corrected estimators			Uncorrected estimators		
	$\tilde{\gamma}_0$	$\tilde{\gamma}_1$	$\tilde{\gamma}_2$	$\tilde{\gamma}_0$	$\tilde{\gamma}_1$	$\tilde{\gamma}_2$
1931–2009	0.1870 (1.16)	0.9534 (3.60)		0.7715 (5.74)	0.3513 (1.90)	
	0.8630 (2.52)	1.0706 (3.40)	−0.0696 (−2.96)	3.0689 (6.01)	0.0258 (0.34)	−0.1945 (−4.90)
1931–1950	−0.1172 (−0.26)	1.8187 (2.21)		0.7655 (2.25)	1.0590 (1.65)	
	0.6444 (0.64)	2.2536 (2.24)	−0.1116 (−1.70)	5.1474 (3.28)	0.1752 (0.76)	−0.3935 (−3.16)
1951–1970	0.6029 (3.95)	0.2896 (1.41)		0.8329 (3.91)	0.0466 (0.54)	
	1.2403 (3.29)	0.3000 (1.32)	−0.0536 (−1.7)	1.7041 (2.66)	0.0260 (0.23)	−0.0736 (−1.50)
1971–1990	0.5464 (2.46)	0.1204 (0.39)		0.7575 (2.57)	−0.1441 (−0.94)	
	0.8502 (1.29)	0.2058 (0.43)	−0.0350 (−0.81)	2.1360 (2.78)	−0.2638 (−2.21)	−0.1158 (−1.99)
1991–2009	−0.3088 (−0.28)	1.6183 (1.02)		0.7279 (1.10)	0.4486 (0.67)	
	0.7095 (0.49)	1.5465 (1.04)	−0.0787 (−0.68)	3.2995 (1.44)	0.1733 (0.55)	−0.1951 (−1.17)

APT outperformed CAPM. Betas are obtained from simultaneous equation system, and a cross-sectional regression of the security return against its β will be used to test the CAPM. They conclude that the beta estimated from the MIMC model by allowing measurement error on the market portfolio does not significantly improve the OLS beta estimate, and MLE estimator does a better job than the OLS and GLS estimators in the cross-sectional regressions because the MLE estimator takes care of the measurement error in beta.

More recently, Jagannathan et al. (2010) and Da et al. (2012) use three-stage cross-sectional regression to correct the errors-in-variables problem from the rolling-window betas. Their empirical findings support CAPM in explaining option-adjusted stock returns at the individual stock level.

100.4.3 Capital Structure

Titman and Wessels (1988), Chang et al. (2009), and Yang et al. (2009) use structure equation models (e.g., LISREL model and MIMIC model) to mitigate the measurement problems of proxy variables when working on capital structure theory. Titman and Wessels (1998) use LISREL method to investigate determinants of capital structure. In the structure equation model, they use 15 indicators associated with 8 latent variables and set 105 restrictions on the coefficient matrix. Empirical results, however, do not support four of eight propositions on the determinants of capital structure. Specifically, their results show that a firm's capital structure is not significantly related to its non-debt tax shields, volatility of earnings, collateral value of assets, and future growth. One possible reason for the poor results is that the indicators used in the empirical study do not adequately reflect the nature of the attributes suggested by financial theory.

Chang et al. (2009) apply a Multiple Indicators and Multiple Causes model (MIMIC) with refined indicators to reexamine Titman and Wessels (1998) work on determinants of capital structure. Chang et al. (2009) examine the seven indicator factors as follows: growth, profitability, collateral value, volatility, non-debt tax shields, uniqueness, and industry. Their empirical results show that the growth is the most influential determinant on capital structure, followed by profitability, and then collateral value. Under a simultaneous cause-effect framework, their seven constructs as determinants of capital structure have significant effects on capital structure decision.

Yang et al. (2009) apply a LISREL model to find determinants of capital structure and stock returns, and estimate the impact of unobservable attributes on capital structure decisions and stock returns. Using leverage ratios and stock returns as two endogenous variables and 11 latent factors as exogenous variables, Yang et al. (2009) find that stock returns, expected growth, uniqueness, asset structure, profitability, and industry classification are the main determinants of capital structure, while leverage, expected growth, profitability, firm value, and liquidity can explain stock returns. In addition, the capital structure and stock return are mutually determined by each other.

Chen et al. (2019) develop a simultaneous determination model to identify the joint determinants of capital structure and stock returns. The structural equation model with confirmatory factor analysis shows that stock return, asset structure, growth rate, industry classification, uniqueness, volatility and financial rating, profitability, government financial policy, and managerial entrenchment are key factors in determining a firm's capital structure. Such results are robust in the MIMIC and two-stage least square methods.

100.4.4 Measurement Error in Investment Equation

Modern q theory, developed by Lucas and Prescott (1971) and Mussa (1977), shows that the shadow value of capital, marginal q , is the firm manager's expectation of the marginal contribution of new capital goods to future profits. Marginal q , therefore, should summarize the effects of all factors relevant to the investment decision. Lucas and Prescott (1971) and Hayashi (1982) show that the equality of marginal q with average q is under the assumptions of constant returns to scale and perfect competition. Because the marginal q is unobservable in the real world, most of empirical studies adopt Lucas and Prescott (1971) and Hayashi's (1982) assumption and use average q instead of marginal q to test q theory. In addition, if financial markets' valuation of the capital will be equal to the manager's valuation, average q should equal an observable value, Tobin's q , defined as the ratio of the market value to the replacement value. Most empirical studies use Tobin's q as a proxy for marginal q to test the q theory of investment. However, their empirical results are inconsistent to the q theory (e.g., Blundell et al. 1992; Fazzari et al. 1988; Gilchrist and Himmelberg 1995; Schaller 1990).⁷

The model introduced by Fazzari et al. (1988) is

$$I_{it}/K_{it} = \eta_i + \beta q_{it}^* + \alpha CF_{it}/K_{it} + u_{it}, \quad (51)$$

where I_{it} represents the investments of firm i at time t , K_{it} is capital stock of firm i at time t , q_{it}^* is the marginal q , CF_{it} is cash flow of firm i at time t , η_i is the firm-specific effect, and u_{it} is the innovation term.

Almeida et al. (2010) show that OLS estimated coefficient of independent variable with measurement error, q_{it}^* , will be biased downward, and OLS estimated coefficient of the independent variable without measurement error, CF_{it}/K_{it} , will be biased upward. Following Eq. (51), if one of independent variables has measurement error and the other independent variables has no measurement error, the asymptotic biases of estimated coefficients can be defined as:

⁷Empirical work in testing association between the investment decision and cash flow shows that cash flow has poor explanation in determining investment decision. In addition to cash flow, output, sales, and internal funds have significant explanation in determining investment decision.

$$\text{plim } \hat{\beta} - \beta = \frac{-\beta\sigma_e^2}{\sigma_{CF/K}^2 - b_{CF/K,q^*} + \sigma_e^2}, \text{ and} \quad (52)$$

$$\text{plim } \hat{\alpha} - \alpha = \beta b_{CF/K,q^*} \left(\frac{\sigma_e^2}{\sigma_e^2 + \sigma_{CF/K}^2 (1 - R_{CF/K,q^*}^2)} \right), \quad (53)$$

in which $\sigma_{CF/K}^2$ is the variance of CF/K , σ_e^2 is the variance of error term between unobserved marginal q and observable average q^* , $b_{CF/K,q^*}$ is the auxiliary regression coefficient of a regressing q^* on CF/K , and $R_{CF/K,q^*}^2$ is the correlation coefficient between q^* and CF/K . We know that $\sigma_{CF/K}^2 - b_{CF/K,q^*} + \sigma_e^2$ is generally positive, so the estimated coefficient of q^* is downward biased. In addition, the direction of bias of estimated coefficient of CF/K will depend on the signs of β and $b_{CF/K,q^*}$. Given that q and cash flow are positively correlated, we can get the conclusion of Almeida et al. (2010) that $\hat{\beta}$ is downward biased and $\hat{\alpha}$ is upward biased.

Erickson and Whited (2000) argue that the measurement error of marginal q can result in different implications in empirical q models. They incorporate an EIV model to reexamine the empirical work done by Fazzari et al. (1988). By using generalized method of moments (GMM), Erickson and Whited (2000) obtain consistent estimators that use the information contained in the third- and higher-order moments of the joint distribution of the observed regression variables. The estimator precision and consistency can be increased by exploiting the information afforded by an excess of moment equations over parameters. Results show that cash flow does not affect a firms' financial decision, even for financially constrained firms, and the q theory is held if measurement error is taken into account.

Almeida et al. (2010) use Monte Carlo simulations and real data to compare the performance of generalized method of moments and instrumental variables approach dealing with measurement error problems in investment equations. In Monte Carlo simulations, they find estimators of GMM proposed by Erickson and Whited (2000) are biased for both mismeasured and well-measured regressors when the data have individual-fixed effects, heteroscedasticity, or no high degree of skewness. In contrast, the instrumental variable method results in fairly unbiased estimators under those same conditions. Almeida et al. (2010) further empirically examine the investment equation introduced by Fazzari et al. (1988) by using GMM and instrumental variable method. Almeida et al. (2010) adopt Biorn's (2000) method using the lags of the variable as instruments in testing the investment equation. Empirical results show that estimators from generalized method of moments are unstable across different specifications and not economically meaningful, while estimators from a simple instrumental method are robust and conform to q theory. Almeida et al. (2010) conclude that instrumental method yields more consistent estimators and support the q theory in the investment equation.

In contrast, Erickson and Whited (2012) compare the ability of three errors-in-variables models, instrumental variables, dynamic panel estimators, and the high-order moment estimators in investment equation. They conclude that all of three

models can perform well under correct specification, while the high-order moment estimators often outperform the instrumental variables and dynamic panel estimators in terms of bias and dispersion. Erickson and Whited (2012) also demonstrate a minimum distance technique to extend the high-order moment estimators used on unbalanced panel data.

100.5 Conclusion

In this chapter, we investigate theoretical issues related to errors-in-variables (EIV) problem, and review how existing EIV estimation methods deal with measurement error problem. We first show how EIV problems affect the coefficients of independent variables in the regression model. We then discuss how classical method, mathematical programming method, grouping method, instrumental variable method, maximum likelihood method, and LISREL method deal with EIV problems. We further investigate how alternative EIV models have been used in empirical finance research. We find that the empirical research of cost of capital, asset pricing, capital structure, and investment equation have used alternative EIV methods to improve the empirical results. Not only can the reader of this chapter understand the important research topics in finance, but also can realize how measurement error problems affect the results of empirical work in such research topics. Finally, we suggest that future empirical studies on finance-related issues should pay more efforts to deal with EIV problems and obtain more robust empirical results.

References

- Ahn, D., J. Conrad, and R.F. Dittmar. 2009. Basis asset. *Review of Financial Studies* 22: 5122–5174.
- Almeida, H., M. Campello, and A.F. Galvao Jr. 2010. Measurement errors in investment equations. *Review of Financial Studies* 23: 3279–3328.
- Anderson, T.W. 1963. The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika* 28: 1–25.
- Ang, A., and J. Chen. 2007. CAPM over the long run: 1926–2001. *Journal of Empirical Finance* 14: 1–40.
- Barnett, V.D. 1967. A note on linear structural relationship when both residual variances are known. *Biometrika* 54: 670–672.
- Bentler, P.M. 1983. Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika* 48: 493–517.
- Bjorn, E. 2000. Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences. *Econometric Reviews* 19: 391–424.
- Black, F., M.C. Jensen, and M. Scholes. 1972. *The capital asset pricing model: Some 1093 empirical tests. In studies in the theory of capital markets*. New York: Praeger.
- Blume, M.E., and I. Friend. 1973. A new look at the capital asset pricing model. *Journal of Finance* 28: 19–33.
- Blundell, R., S. Bond, M. Devereux, and F. Schiantarelli. 1992. Investment and Tobin's Q: Evidence from company panel data. *Journal of Econometrics* 51: 233–257.

- Brennan, M.J. 1970. Taxes, market valuation and corporate financial policy. *National Tax Journal* 23: 417–427.
- . 1979. The pricing of contingent claims in discrete time models. *Journal of Finance* 34: 53–68.
- Brennan, M.J., and E.S. Schwartz. 1977. Convertible bonds: Valuation and optimal strategies for call and conversion. *Journal of Finance* 32: 1699–1715.
- Brown, S.J., and J.B. Warner. 1980. Measuring security price performance. *Journal of Financial Economics* 8: 205–258.
- Chang, C., A.C. Lee, and C.F. Lee. 2009. Determinants of capital structure choice: A structural equation modeling approach. *Quarterly Review of Economics and Finance* 49: 197–213.
- Chen, H. Y. 2011. *Momentum strategies, dividend policy, and asset pricing test*. Ph.D. dissertation, State University of New Jersey, Rutgers.
- Chen, H.Y., C.F. Lee, and T. Tai. 2019. The joint determinants of capital structure and stock rate of return: A LISREL model approach. *Review of Pacific Basin Financial Markets and Policies* 22: 1950013.
- Cheng, P.L., and R.R. Grauer. 1980. An alternative test of the capital asset pricing model. *American Economic Review* 70: 660–671.
- Clutton-Brock, M. 1967. Likelihood distributions for estimating functions when both variables are subject to error. *Technometrics* 9: 261–269.
- Cochran, W.G. 1970. Some effects of errors of measurement on multiple correlation. *Journal of the American Statistical Association* 65: 22–34.
- Da, Z., R.J. Guo, and R. Jagannathan. 2012. CAPM for estimating the cost of equity capital: Interpreting the empirical evidence. *Journal of Financial Economics* 103: 204–220.
- Davis, P. 2010. *A firm-level test of the CAPM*. Working paper, Rutgers University.
- Deming, W.E. 1943. *Statistical adjustment of data*. New York: Wiley.
- Diacogiannis, G., and D. Feldman. 2011. *Linear beta pricing with inefficient benchmarks*. Working paper, University of Piraeus.
- Durbin, J. 1954. Errors in variables. *Review of the International Statistical Institute* 22: 23–32.
- Erickson, T., and T. Whited. 2000. Measurement error and the relationship between investment and q. *Journal of Political Economy* 108: 1027–1057.
- . 2002. Two-step GMM estimation of the errors-in-variables model using higher-order moments. *Econometric Theory* 18: 776–799.
- . 2012. Treating measurement error in Tobin's Q. *Review of Financial Studies* 25: 1286–1329.
- Fabozzi, F.J., and J.C. Francis. 1978. Beta as a random coefficient. *Journal of Financial and Quantitative Analysis* 13: 101–116.
- Fama, E.F., and K.R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47: 427–466.
- Fama, E.F., and J.D. MacBeth. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 81: 607–636.
- Fazzari, S., R.G. Hubbard, and B. Petersen. 1988. Financing constraints and corporate investment. *Brookings Papers on Economic Activity* 1: 141–195.
- Geweke, J., and G. Zhou. 1996. Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies* 9: 557–587.
- Gibbons, M.R. 1982. Multivariate tests of financial models: A new approach. *Journal of Financial Economics* 10: 3–27.
- Gibbons, M.R., and W. Ferson. 1985. Testing asset pricing models with changing expectations and an unobservable market portfolio. *Journal of Financial Economics* 14: 217–236.
- Gilchrist, S., and C.P. Himmelberg. 1995. Evidence on the role of cash flow for investment. *Journal of Monetary Economics* 36: 541–572.
- Goldberger, A.S. 1972. Structural equation methods in the social sciences. *Econometrica* 40: 979–1001.
- Green, R. 1986. Benchmark portfolio inefficiency and deviations from the security market line. *Journal of Finance* 41: 295–312.
- Griliches, Z., and J.A. Hausman. 1986. Errors in variables in panel data. *Journal of Econometrics* 31: 93–118.

- Guay, W., S.P. Kothari, and S. Shu. 2011. Properties of implied cost of capital using analysts' forecasts. *Australian Journal of Management* 36: 125–149.
- Hayashi, F. 1982. Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica* 50: 213–224.
- Higgins, R.C. 1974. Growth, dividend policy and capital cost in the electric utility industry. *Journal of Finance* 29: 1189–1201.
- Jagannathan, R., and Z. Wang. 1996. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51: 3–53.
- . 1998a. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51: 3–53.
- . 1998b. An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *Journal of Finance* 53: 1285–1309.
- Jagannathan, R., G. Skoulakis, and Z. Wang. 2009. The analysis of the cross section of security returns. In *Handbook of financial econometrics*, ed. Y. Aït-Sahalia and L. Hansen, vol. 2, 73–134. Amsterdam: North-Holland.
- Jagannathan, R., S. Kim, and G. Skoulakis. 2010. *Revisiting the errors in variables problem in studying the cross section of stock returns*. Working paper, Northwestern University.
- Jöreskog, K.G., and A.S. Goldberger. 1975. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association* 70: 631–639.
- Jöreskog, K.G., and D. Sörbom. 1981. *LISREL V: Analysis of linear structural relationships by the method of maximizing likelihood, user's guide*. Chicago: National Educational Resources.
- . 1989. *LISREL 7: A guide to the program and applications*. Chicago: SPSS.
- Kendall, M.G., and A. Stuart. 1961. *The advanced theory of statistics*. Vol. II. London: Griffin.
- Kiefer, J. 1964. Review of Kendall and Stuart's advanced theory of statistics – II. *Annals of Mathematical Statistics* 35: 1371–1380.
- Kim, D. 1995. The errors in the variables problem in the cross-section of expected stock returns. *Journal of Finance* 50: 1605–1634.
- . 1997. A reexamination of firm size, book-to-market, and earnings price in the cross-section of expected stock returns. *Journal of Financial and Quantitative Analysis* 32: 463–489.
- . 2010. Issues related to the errors-in-variables problems in asset pricing tests. In *Handbook of quantitative finance and risk management*, ed. C.F. Lee, A.C. Lee, and J. Lee. Springer.
- Lee, Cheng F. 1973. *Errors-in-variables estimation procedures with applications to a capital asset pricing model*. Ph.D. dissertation, State University of New York at Buffalo.
- . 1977. Performance measure, systematic risk and errors-in-variable estimation method. *Journal of Economics and Business* 29: 122–127.
- . 1984. Random coefficient and errors-in-variables models for beta estimates: Methods and application. *Journal of Business Research* 12: 505–516.
- Lee, C.F., and S.N. Chen. 1979. A random coefficient model for reexamining risk decomposition method and risk-return relationship test. *Financial Review* 14: 65–65.
- Lee, Cheng F., and F.C. Jen. 1978. Effects of measurement errors on systematic risk and performance measure of a portfolio. *Journal of Financial and Quantitative Analysis* 13: 299–312.
- Lee, Cheng F., and K. C. Wei. 1984. *Multi-factor, multi-indicator approach to asset pricing: Methods and empirical evidence*. Working paper, University of Illinois at Urbana-Champaign.
- Lee, Cheng F., and C.C. Wu. 1989. Using Zellner's errors-in-variables model to reexamine MM's valuation model for the electric utility industry. *Advances in Financial Planning and Forecasting* 3: 63–73.
- Lee, C.F., H.Y. Chen, and J. Lee. 2019. *Financial econometrics, mathematics and statistics: theory, method and application*. New York: Springer.
- Lewbel, A. 1997. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica* 65: 1201–1213.
- . 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics* 30: 67–80.

- Lintner, J. 1965. The valuation of risky assets and the selection of risky investments in stock portfolio and capital budgets. *Review of Economics and Statistics* 47: 13–37.
- Litzenberger, R., and K. Ramaswamy. 1979. The effects of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics* 7: 163–195.
- Lucas, R.E., Jr., and E.C. Prescott. 1971. Investment under uncertainty. *Econometrica* 39: 659–681.
- MacKinlay, A.C., and M.P. Richardson. 1991. Using generalized method of moments to test mean-variance efficiency. *Journal of Finance* 46: 511–527.
- Maddala, G.S., and M. Nimalendran. 1996. Error-in-variables problems in financial models. In *Handbook of statistics*, ed. G.S. Maddala and C.R. Rao. New York: Elsevier Science.
- McCulloch, R., and P.E. Rossi. 1991. A Bayesian approach to testing the arbitrage pricing theory. *Journal of Econometrics* 49: 141–168.
- Miller, M.H., and F. Modigliani. 1966. Some estimates of the cost of capital to the electric utility industry, 1954–57. *American Economic Review* 56: 333–391.
- Mossin, J. 1966. Equilibrium in a capital asset market. *Econometrica* 34: 768–783.
- Mussa, M. 1977. External and internal adjustment costs and the theory of aggregate and firm investment. *Economica* 44: 163–178.
- Ortiz-Molina, H., and G.M. Phillips. 2014. Real asset illiquidity and the cost of capital. *Journal of Financial and Quantitative Analysis* 49: 1–32.
- Pastor, L., M. Sinha, and B. Swaminathan. 2008. Estimating the intertemporal risk-return tradeoff using the implied cost of capital. *Journal of Finance* 63: 2859–2897.
- Roll, R. 1969. Bias in fitting the Sharpe model to time-series data. *Journal of Financial and Quantitative Analysis* 4: 271–289.
- . 1977. A critique of the asset pricing theory's tests: Part 1. On past and potential testability of the theory. *Journal of Financial Economics* 4: 129–176.
- Roll, R., and S.A. Ross. 1994. On the cross-sectional relation between expected returns and betas. *Journal of Finance* 49: 101–121.
- Schaller, H. 1990. A re-examination of the q theory of investment using US firm data. *Journal of Applied Econometrics* 5: 309–325.
- Shanken, J. 1992. On the estimation of beta pricing models. *Review of Financial Studies* 5: 1–33.
- Sharpe, W.F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- Stapleton, D.C. 1978. Analyzing political participation data with a MIMIC model. In *Sociological methodology*, ed. K.F. Schuessler. San Francisco: Jossey-Bass.
- Theil, H. 1971. *Principles of econometrics*. Toronto: Wiley.
- Titman, S., and R. Wessels. 1988. The determinants of capital structure choice. *Journal of Finance* 43: 1–19.
- Wald, A. 1940. The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11: 284–300.
- Wei, K. C. 1984. *The arbitrage pricing theory versus the generalized intertemporal capital asset pricing model: Theory and empirical evidence*. Ph.D. dissertation, University of Illinois at Urbana-Champaign.
- Yang, C.C., C.F. Lee, Y.X. Gu, and Y.W. Lee. 2009. Co-determination of capital structure and stock returns – a LISREL approach: An empirical test of Taiwan stock markets. *Quarterly Review of Economics and Finance* 50: 222–233.
- York, D. 1966. Least-squares fitting of a straight line. *Canadian Journal of Physics* 44: 1079–1086.
- Zellner, A. 1970. Estimation of regression relationships containing unobservable variables. *International Economic Review* 11: 441–454.



Optimal Payout Ratio under Uncertainty and the Flexibility Hypothesis: Theory and Empirical Evidence

101

Cheng-Few Lee, Manak C. Gupta, Hong-Yi Chen, and Alice C. Lee

Abstract

Following the dividend flexibility hypothesis used by DeAngelo and DeAngelo (J Financ Econ 81, 227–254, 2006), Blau and Fuller (J Corp Financ 14:133–152, 2008), and others, we theoretically extend the proposition of DeAngelo and DeAngelo's (J Financ Econ 81, 227–254, 2006) optimal payout policy in terms of the flexibility dividend hypothesis. We also introduce growth rate, systematic risk, and total risk variables into the theoretical model. In addition, based upon Lee and Lee (2021), we discuss the implication of the existence of optimal payout ratio in financial analysis and decision for a company.

To test the theoretical results derived in this chapter, we use data collected in the USA from 1969 to 2009 to investigate the impact of growth rate, systematic risk, and total risk on the optimal payout ratio in terms of the fixed-effect model. We find that based on flexibility considerations, a company will reduce its payout when the growth rate increases. In addition, we find that a nonlinear relationship exists between the payout ratio and the risk. In other words, the relationship between the payout ratio and risk is negative (or positive) when the growth rate is higher (or lower) than the rate of return on total assets. Our theoretical model and

C.-F. Lee

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA

e-mail: lee@business.rutgers.edu

M. C. Gupta

Temple University, Philadelphia, PA, USA

e-mail: mcgupta@temple.edu

H.-Y. Chen (✉)

Department of Finance, National Chengchi University, Taipei, Taiwan

e-mail: fnhchen@nccu.edu.tw

A. C. Lee

Center for PBBEF Research, Phoenix, AZ, USA

empirical results can therefore be used to identify whether flexibility or the free cash flow hypothesis should be used to determine the dividend policy.

Keywords

Dividends · Flexibility hypothesis · Payout policy · Fixed-effects model · Financing policy · Investment policy

101.1 Introduction

Corporate dividend policy has long engaged the attention of financial economists, dating back to the irrelevance theorem of Miller and Modigliani (1961; M&M hereafter), in which they state that a rational and perfect economic environment is free of illusion. Since then, their rather controversial findings have been challenged and tested by weakening the assumptions or introducing imperfections into the analysis. For example, the signaling models developed by Bhattacharya (1979), Miller and Rock (1985), and John and Williams (1985), and the free cash flow hypothesis proposed by Easterbrook (1984) and Jensen (1986) are the two well-known propositions challenging M&M's dividend irrelevance theorem; however, empirical studies examining signaling and free cash flow hypotheses have yielded mixed results.

DeAngelo and DeAngelo (2006) reexamine the irrelevance of the M&M dividend irrelevance theorem by allowing partial dividend payout. They argue that the original M&M (1961) irrelevance result is due to the fact that they consider either paying out all earnings or not paying any of them at all. Therefore, payout policy might be relevant if partial payout is allowed. In other words, DeAngelo and DeAngelo (2006) use dividend flexibility hypothesis to show their dividend relevance results. Underlying the flexibility hypothesis, the current chapter (1) develops a theoretical model to support the proposition of DeAngelo and DeAngelo's (2006) optimal payout policy when the partial payout is allowed; (2) reconciles the dispute between free cash flow hypothesis and flexibility hypothesis in dividend policy literature; and (3) performs empirical tests in terms of theoretical results derived in this chapter.

First of all, following DeAngelo and DeAngelo (2006), we develop a dynamic model allowing firms to hold some amount of cash into a positive NPV project for the reason of financial flexibility. Under the assumption of stochastic rate of return and the dividend flexibility hypothesis, we carry out the optimization procedure to maximize firm value, and the final expression of the optimal dividend policy of the firm is thus derived. The model is comprehensive and allows structural analysis of different variables that could be relevant for corporate dividend policy. For example, the model incorporates the rate of return on assets, growth rate, and risk (systematic, firm-specific, and total risk) to name a few variables that could affect corporate dividend decisions. Comparative statics provide insights into the effect of each of these parameters on corporate dividend policy, and this is followed by an analysis of the interaction effects of these variables on corporate dividend policy.

Second, the implications of the optimization results are explained. Our results show that the relationship between the optimal payout ratio and the growth rate is negative in general. We investigate the separate and then the combined effects of market-dependent and market-independent components of risk on the optimal dividend policy. We perform comparative static analyses of the relationships between the payout ratio and (1) change in total risk, (2) change in systematic risk, (3) simultaneous changes in both total risk and systematic risk, and (4) no change in risk. We examine in detail the effects of variations in the profitability rate, its distribution parameters, and their dynamic behavior on the optimal dividend policy of the firm. The theoretical relationship between the payout ratio and the growth ratio implies that high growth firms need to reduce the payout ratio and retain more earnings to build up “precautionary reserves” for flexibility considerations, but low growth firms are likely to be more mature and already build up their reserves. More importantly, the relationship between the payout ratio and the risks reflects different dividend policies at high growth firms and low growth firms. With higher risk, the costs of external funds increase; therefore, to optimize shareholders’ wealth, high growth firms tend to reduce their payouts and keep more relatively low-cost funds to sustain their high growth, whereas low growth firms tend to pay more dividends and reduce the risk for shareholders.

Third, our theoretical model and its implications lead us to three testable hypotheses. Although a large and growing body of empirical research on the optimal dividend payout policy has emerged, none of it has a solid theoretical model to support findings or introduces a nonlinear structured model on payout policy.¹ We here try to empirically examine three hypotheses derived from our theoretical optimal payout model. Using data collected in the USA from 1969 to 2009, we analyze 28,333 dividend-paying firm-years. Our empirical results show that firms’ payout ratios are negatively related to firms’ growth and a negative (or positive) relationship exists between firms’ risks and the dividend payout ratios among firms with higher (or lower) growth rates relative to their rate of return on assets. Our empirical results are consistent with our theoretical model under the dividend flexibility hypothesis. We also find that growth and risk interact in explaining the payout ratio, indicating that the payout ratio is not linearly related to the growth rate or to the risk of the firm. Furthermore, we implement the moving estimates process to find the empirical breakpoint of the structural change for the relationship between the payout ratio and risks and confirm that the empirical breakpoint is not different from our theoretical breakpoint.

The primary contributions of this chapter are our theoretical derivation of an optimal payout ratio under the dividend flexibility hypothesis and our demonstration

¹For example, Rozeff (1982), Jagannathan et al. (2000), Grullon et al. (2002), Aivazian et al. (2003), and Blau and Fuller (2008).

of a negative but nonlinear relationship between the payout ratio and the growth rate. More importantly, we theoretically and empirically locate a structural change point for the relationship between the payout ratio and the growth rate. Rozeff (1982), Aivazian et al. (2003), Blau and Fuller (2008), and others conclude that a firm's risk and optimal payout are negatively related. Contrary to these conclusions and generally held beliefs, the dynamic optimization model developed here shows that the optimal payout is not necessarily negatively related to risk but implies that payout policies differ in high growth firms and low growth firms. High growth firms pay out their dividends for flexibility considerations, but low growth firms pay out their dividends to reduce their free cash flow problem. In addition, this chapter also shows that the relationship between firms' payouts and their growth rates (or risks) can be affected by their risks (or growth rates). Contrary to earlier studies, the theoretical model developed here shows that the optimal payout ratio is not linearly related to the growth rate or to the risk of the firm, and whether the rate of return on assets is higher or lower than the growth rate has significant effect on the relationships between these variables. The results are borne out by rather extensive empirical research done in this chapter based on 40 years of data collected in the USA from 1969 to 2009.

The stochastic dynamic optimization model developed here challenges many commonly held beliefs and results obtained from earlier studies. It provides more meaningful relationships among total risk and its separate components (systematic and firm-specific), growth rate, rate of return on assets, and optimal dividend policy. Our results represent an advance in corporate finance literature because all the important variables included in this study were never simultaneously included in those earlier studies, and more importantly, the interaction effects among these variables were neither recognized nor fully appreciated. We believe that this study is the first of its kind to provide a theoretical basis for relating dividend policies with firm risks, growth rates, and return on assets (ROAs) in an interrelated fashion.

The remainder of this chapter is organized as follows. Section 101.2 contains a review of the literature on dividend-relevant theories and the findings of empirical work on dividend policy. In Sect. 101.3, we lay out a dynamic model used in subsequent sections to examine the existence, or nonexistence, of an optimal dividend policy. Section 101.4 provides the final expression of the optimal dividend policy of the firm derived from the optimization procedure to maximize firm value. In Sect. 101.5, we present both a detailed form and an approximated form of the relationship between the optimal dividend payout ratio and the growth rate. Section 101.6 includes a discussion of the effects of market-dependent and market-independent components of risk on the optimal dividend policy; Section 101.7 includes empirical evidence supporting the model and implications in previous sections; Sect. 101.9 discusses the potential interaction among dividend, investment, financing, and production policies, and the conclusion appears in Sect. 101.10.

101.2 Review of the Literature

Corporate dividend policy has puzzled financial economists, dating back to the dividend irrelevance theorem proposed by M&M (1961). Since then, their controversial findings have been challenged and tested by weakening the assumptions or introducing imperfections into the analysis. The signaling models developed by Bhattacharya (1979), Miller and Rock (1985), and John and Williams (1985) suggest that because of the asymmetric information between managers and shareholders, managers use dividends as a signal to release private information to the market; however, empirical studies examining the signaling hypothesis have yielded mixed results. Nissim and Ziv (2001), Brook et al. (1998), Bernheim and Wantz (1995), Kao and Wu (1994), and Healy and Palepu (1988) support the signaling (asymmetric information) hypothesis by finding a positive association between dividend increases and future profitability. Kalay and Lowenstein (1986) and Asquith and Mullins (1983) find that dividend changes are positively associated with stock returns in the days surrounding dividend announcements, and Sasson and Kolodny (1976) identify that a positive association between the payout ratio and average rates of return. Studies by Benartzi et al. (1997), DeAngelo et al. (1996), and Grullon et al. (2002), however, reveal no support for the hypothesized relationship between dividend changes and future profitability.

Using agency cost theory, Easterbrook (1984) and Jensen (1986) propose the free cash flow hypothesis, arguing that because managers cannot credibly pre-commit to shareholders, they will not invest excess cash in negative-NPV projects. Dividend changes may convey information about how the firm will use future cash flow. Again, the results of empirical studies have been mixed at best. Several researchers, including Agrawal and Jayaraman (1994), Jensen et al. (1992), and Lang and Litzenberger (1989) find positive supports for the agency cost hypothesis, but others find no support for this hypothesis, for example, Howe et al. (1992), Denis et al. (1994), and Yoon and Starks (1995).

DeAngelo and DeAngelo (2006) argue that M&M's (1961) dividend irrelevance occurs because they consider either paying out all earnings or not paying any of them at all; therefore, payout policy might have impact on firm value if partial payout is allowed. In other words, DeAngelo and DeAngelo (2006) use the dividend flexibility hypothesis to show their dividend relevance results. Blau and Fuller (2008) have theoretically and empirically shown that the dividend flexibility hypothesis is indeed a reasonable dividend policy. In addition, several empirical studies show evidence of firms preferring financial flexibility, for example, Lie (2005), DeAngelo et al. (2006), Denis and Osobov, (2008), and Gabudean (2007).

Besides focusing on the relevance of dividend policy, a growing body of literature deals with the determinants of optimal dividend payout policy. For example, Rozeff (1982) shows that the optimal dividend payout is related to the fraction of insider holdings, the growth of the firm, and the firm's beta coefficient. He also finds evidence that the optimal dividend payout is negatively correlated to

beta risk, and supporting that beta risk reflects the leverage level of a firm. Jagannathan et al. (2000) empirically show that operation risk is negatively related to the propensity to increase payouts. Grullon et al. (2002) show that dividend changes are related to the change in the growth rate and the change in the rate of return on assets. They also find that dividend increases should be associated with subsequent declines in profitability and risk. Aivazian et al. (2003) examine eight emerging markets and show that, similar to the US firms, dividend policies in emerging markets can also be explained by profitability, debt, and the market-to-book ratio. None of the foregoing scholars, however, has a solid theoretical model to support their finding.

We use these authors' works as a springboard to deal with the flexibility hypothesis and develop the theoretical underpinnings and a stochastic dynamic optimization model determining the conditions for an optimal dividend policy. We believe that ours is the first attempt ever made to include growth rate, risks, and more importantly, the rate of return and its distribution parameters in a single model deriving the conditions for an optimal dividend policy; furthermore, this chapter empirically tests the findings and conclusions of the theoretical model developed here, using an extensive data set involving 28,333 dividend-paying firm-years from 1969 to 2009. The results of the empirical tests confirm and validate the findings of the theoretical model and provide new insights in the determination of optimal dividend policy, the influence of each explanatory variable listed above, and more importantly, their interaction effects on the corporate dividend policy. Notably, some of these results run counter to categorically stated conclusions reached in earlier studies.

101.3 The Model

We develop the dividend policy model under the assumptions that the capital markets represent the closest approximation to the economists' ideal of a perfect market – zero transaction costs, rational behavior on the part of investors, and the absence of tax differentials between dividends and capital gains. It is assumed that the firm is not restricted to financing its growth only by retained earnings, and that its rate of return, $\tilde{r}(t)$, is a nonstationary random variable, normally distributed with mean, μ , and variance, $\sigma(t)^2$.

Let $A(0)$ represent the initial assets of the firm and h be the growth rate. Then the earnings of this firm are given by Eq. (1), which is

$$\tilde{x}(t) = \tilde{r}(t)A(0)e^{ht}, \quad (1)$$

where $\tilde{x}(t)$ represents the earnings of the firm, and the tilde ($\sim$) denotes its random character.

Based upon DeAngelo and DeAngelo's (2006) assumption, we here allow that the firm partially pay out its earnings and retain a certain amount of earnings in

support of its growth.² The retained earnings of the firm, $y(t)$, can therefore be expressed as follows:

$$y(t) = \tilde{x}(t) - m(t)\tilde{d}(t), \quad (2)$$

where $\tilde{d}(t)$ is the dividends per share, and $m(t)$ is the total number of shares outstanding at time t .

Equation (2) further indicates that the focus of the firm's decision-making is on retained earnings, which implies that dividend $\tilde{d}(t)$ also becomes a random variable. The growth of a firm can be financed by retained earnings or by issuing new equity.

The new equity raised by the firm at time t can be defined as follows:

$$e(t) = \delta p(t)\dot{m}(t), \quad (3)$$

where

$p(t)$ = price per share

$\dot{m}(t) = dm(t)/dt$

δ = degree of market perfection, $0 < \delta \leq 1$

The value of δ equal to one indicates that new shares can be sold by the firm at current market prices.

From Eqs. (1, 2, and 3) investment in period t is the sum of retained earnings and funds raised by new equity, so the investment in period t can be written as follows:

$$hA(0)e^{ht} = \tilde{x}(t) - m(t)\tilde{d}(t) + \delta\dot{m}(t)p(t). \quad (4)$$

This implies that

$$\tilde{d}(t) = \{[\tilde{r}(t) - h]A(0)e^{ht} + \delta\dot{m}(t)p(t)\}/m(t), \quad (5)$$

and the mean and variance of the dividends per share can be expressed as follows:

$$\begin{aligned} E[\tilde{d}(t)] &= \{[\mu - h]A(0)e^{ht} + \delta\dot{m}(t)p(t)\}/m(t) \\ Var[\tilde{d}(t)] &= A(0)^2\sigma(t)^2e^{2ht}/m^2(t) \end{aligned} \quad (6)$$

²DeAngelo and DeAngelo (2006) carefully explain why partial payout is important to obtain an optimal ratio under perfect markets. In addition, they also argue that partial payout is important to avoid the suboptimal solution for optimal dividend policy.

Also, let us postulate an exponential utility function of the following form³:

$$U[\tilde{d}(t)] = -e^{-\alpha \tilde{d}(t)}, \quad (7)$$

where $\alpha > 0$

Following the moment generating function, we have

$$E\left(-e^{-\alpha \tilde{d}(t)}\right) = -e^{-\alpha E[\tilde{d}(t)] + \frac{\alpha^2}{2} \text{Var}[\tilde{d}(t)]}, \quad (8)$$

where $\tilde{d}(t)$ is the certainty equivalent value of $\tilde{d}(t)$.⁴

From Eqs. (6) and (8), the certainty equivalent dividend stream can be written as

$$\bar{d}(t) = \frac{(\mu - h)A(0)e^{th} + \delta \dot{m}(t)p(t)}{m(t)} - \frac{\alpha' A(0)^2 \sigma(t)^2 e^{2th}}{m(t)^2}, \quad (9)$$

where $\alpha' = \alpha/2$. Therefore, taking advantage of exponential utility, we can obtain a risk adjusted dividend stream. Furthermore, $\bar{d}(t)$ will reduce to the certainty case, if we assume $\sigma(t)^2 = 0$.

In accordance with the capital asset pricing theory developed by Sharpe (1964), Lintner (1963), and Mossin (1966), the total risk can be decomposed into systematic risk and unsystematic risk; that is, $\tilde{r}(t)$ can be defined as follows:

$$\tilde{r}(t) = a + b\tilde{I}(t) + \tilde{\varepsilon}(t), \quad (10)$$

where $\tilde{I}(t)$ is the market index; $\tilde{\varepsilon}(t) \sim N(0, \sigma_\varepsilon^2)$; a and b are regression parameters; and $\text{Var}(b\tilde{I}(t))$ and $\text{Var}(\tilde{\varepsilon}(t))$ represent the systematic and unsystematic risk, respectively.

Following Eq. (10), Eq. (6) can be rewritten as

$$\begin{aligned} E[\tilde{d}(t)] &= [(a + b\tilde{I} - h)A(0)e^{ht} + \delta \dot{m}(t)p(t)]/m(t) \\ \text{Var}[\tilde{d}(t)] &= A(0)^2 [b^2 \text{Var}(\tilde{I}(t)) + \text{Var}(\varepsilon(t))] e^{2th}/m(t)^2, \\ &= A(0)^2 [\rho(t)^2 \sigma(t)^2 + (1 - \rho(t)^2) \sigma(t)^2] e^{2th}/m(t)^2 \end{aligned} \quad (11)$$

³Pratt (1964) provides a detailed analysis of the various utility functions. Exponential, hyperbolic, and quadratic forms have been variously used in the literature; but the first two seem to have preference over the quadratic form because the latter has the undesirable property that it ultimately turns downwards.

⁴From the moment generating function discussed in Hogg and Craig (2004), we know that $E(-e^{ty}) = -e^{tE(y) + \frac{1}{2}t^2 \text{Var}(y)}$. Let $t = -\alpha$, then the right hand side of (8) is easily obtained.

where

$\rho(t)$ = the correlation coefficient between $\tilde{r}(t)$ and $\tilde{I}$
 a = market-independent component of the firm's rate of return
 $b\tilde{I}$ = market-dependent component of the firm's rate of return
 $\rho(t)^2\sigma(t)^2$ = nondiversifiable risk
 $(1 - \rho(t)^2)\sigma(t)^2$ = diversifiable risk

The unsystematic risk usually can be diversified away by the investors,⁵ so the certainty equivalent value in Eq. (9) should be revised as

$$\hat{d}'(t) = \frac{(a + b\tilde{I} - h)A(o)e^{th} + \delta\dot{m}(t)p(t)}{m(t)} - \frac{\alpha'A(o)^2\rho(t)^2\sigma(t)^2e^{2th}}{m(t)^2}. \quad (12)$$

Following Lintner (1962), we observe that the stock price should equal the present value of this certainty equivalent dividend stream discounted at a riskless rate of return. Therefore,

$$p(0) = \int_0^T \hat{d}'(t)e^{-kt} dt, \quad (13)$$

where

$p(0)$ = the stock price at $t = 0$
 k = the risk free rate of return
 T = the planning horizon

This model will be used in subsequent sections to find the functional form of $m(t)$ and optimize the payout ratio. The formulation of our model is different from that of M&M (1961), Gordon (1962), Lerner and Carleton (1966a), and Lintner (1964). For example, in contrast to our model, M&M consider neither the non-stationarity of the firm's rate of return in their model nor explicitly incorporated uncertainty in their valuation model. Their models are also essentially static and would not permit an extensive analysis of the dynamic process of moving from one equilibrium state to another; furthermore, the formulation of our model is different from those who propose to capitalize the market-dependent and market-independent components of the uncertain stream of earnings at the risky and riskless rates, respectively.⁶ Instead, we view the market value of a firm as the present value of certainty equivalents of random future receipts. In the next section,

⁵See Lintner (1965), Mossin (1966), and Sharpe (1964).

⁶See Brennan (1973).

we carry out the optimization of Eq. (13) and derive the final expression for the optimal payout ratio.⁷

101.4 Optimum Dividend Policy

Based upon the evaluation model developed in the previous section, in this section we will derive an optimal dividend payout ratio.

Substituting Eq. (12) into Eq. (13), we obtain

$$p(0) = \int_0^T \left[\frac{(A + b\bar{I} - h)A(0)e^{th} + \delta\dot{m}(t)p(t)}{m(t)} - \frac{\alpha'A(0)^2\rho(t)^2\sigma(t)^2e^{2th}}{m(t)^2} \right] e^{-kt} dt. \quad (14)$$

To maximize Eq. (14), we observe that

$$p(t) = \int_t^T \hat{d}'(s)e^{-k(s-t)} ds = e^{kt} \int_t^T \hat{d}'(s)e^{-ks} ds, \quad (15)$$

where s = the proxy of time in the integration

From Eq. (15), we can formulate a differential equation as

$$\frac{dp(t)}{d(t)} = \dot{p}(t) = kp(t) - \hat{d}'(t). \quad (16)$$

Substituting Eq. (12) into Eq. (16), we obtain the differential equation

$$\dot{p}(t) + \left[\frac{\delta\dot{m}(t)}{m(t)} - k \right] p(t) = -G(t), \quad (17)$$

where

$$G(t) = \frac{(a + b\bar{I} - h)A(0)e^{th}}{m(t)} - \frac{\alpha'A(o)^2\rho(t)^2\sigma(t)^2e^{2th}}{m(t)^2}. \quad (18)$$

Following Kreyszig (2010) and Lee and Shi (2010), we solve the differential Eq. (17) and obtain the solution as indicated in Eq. (19).

⁷For further explanation of the optimization of the deterministic and stochastic control models and their applications to economic problems, please see Aoki (1967), Bellman (1990 and 2003), and Intriligator (2002).

$$p(t) = \frac{e^{kt}}{m(t)^\delta} \int_t^T G(s) m(s)^\delta e^{-ks} ds. \quad (19)$$

Then, Eq. (20) can be obtained from Eq. (18) and Eq. (19), implying that the initial value of a stock can be expressed as the summation of present values of its earnings stream adjusted by the risk taken by the firm.

$$p(0) = \frac{1}{m(0)^\delta} \int_0^T \left\{ (a + b\bar{l} - h) A(0) e^{th} m(t)^{\delta-1} - \alpha' A(0)^2 \rho(t)^2 \sigma(t)^2 e^{2th} m(t)^{\delta-2} \right\} e^{-kt} dt. \quad (20)$$

To maximize firm value, the Euler-Lagrange condition for the optimization of $p(o)$ is given by Eq. (21),⁸

$$(\delta - 1) (a + b\bar{l} - h) A(0) e^{th} m(t)^{\delta-2} - \alpha' A(0)^2 \rho(t)^2 \sigma(t)^2 e^{2th} m(t)^{\delta-3} (\delta - 2) = 0. \quad (21)$$

Therefore, the optimal shares outstanding at time t can be derived.

$$m(t) = \frac{(2 - \delta) \alpha' A(0) e^{th} \rho(t)^2 \sigma(t)^2}{(1 - \delta) (a + b\bar{l} - h)}. \quad (22)$$

From Eqs. (18), (19), and (22), we can obtain the maximized stock value

$$p(t) = \frac{(a + b\bar{l} - h)^2 (1 - \delta) e^{kt - th\delta} \int_t^T e^{\delta hs - ks} (\rho(s) \sigma(s))^{2\delta-2} ds}{\alpha' (2 - \delta)^2 \rho(t)^{2\delta} \sigma(t)^{2\delta}}. \quad (23)$$

From Eq. (22), we also obtain the optimal number of shares of new equity issued at time t

$$\dot{m}(t) = \frac{\left\{ h(2 - \delta) \alpha' A(0) e^{th} \rho(t)^2 \sigma(t)^2 + (2 - \delta) \alpha' A(0) e^{th} \left[\rho(t)^2 \dot{\sigma}(t)^2 + \sigma(t)^2 \dot{\rho}(t)^2 \right] \right\}}{(1 - \delta) (a + b\bar{l} - h)}. \quad (24)$$

From Eq. (23) and Eq. (24), we have the amount generated from issuing new equity

⁸For the derivation of Eq. (21), please refer to the [Appendix B](#).

$$\dot{m}(t)p(t) = \frac{(a + b\bar{l} - h)e^{kt - (\delta-1)th}A(0) \left(h\rho(t)^2\sigma(t)^2 + \rho(t)^2\dot{\sigma}(t)^2 + \sigma(t)^2\dot{\rho}(t)^2 \right) \int_t^T e^{s(\delta h - k)} (\rho(s)\sigma(s))^{2\delta-2} ds}{(2 - \delta)\rho(t)^{2\delta}\sigma(t)^{2\delta}}. \quad (25)$$

From Eq. (5) and Eq. (25), we can obtain $\bar{D}(t) = m(t)\bar{d}(t)$. From Eq. (1) and Eq. (10), we can obtain $\bar{x}(t) = (a + b\bar{l})A(0)e^{ht}$. When δ approaches unity, we can derive the optimal payout ratio as

$$\frac{\bar{D}(t)}{\bar{x}(t)} = \frac{(a + b\bar{l} - h)}{(a + b\bar{l})} \left\{ 1 + \frac{[e^{(h-k)(T-t)} - 1]}{(h - k)} \left(h + \frac{\dot{\sigma}(t)^2}{\sigma(t)^2} + \frac{\dot{\rho}(t)^2}{\rho(t)^2} \right) \right\}. \quad (26)$$

Eq. (26) implies an optimal payout ratio when we use an exponential utility function to derive the stochastic dynamic dividend policy model. This result does not necessarily imply that the dividend policy results derived by M&M (1961) are false because we allowed free cash flow to be paid out partially as assumed by DeAngelo and DeAngelo (2006) instead of paying out all free cash flows as assumed by M&M (1961).

In the following section, we use Eq. (26) to explore the implications of the stochasticity, the stationarity (in the strict sense), and the nonstationarity of the firm's rate of return for its dividend policy.⁹ We also investigate in detail the differential effects of variations in the systematic and unsystematic risk components of the firm's stream of earnings on the dynamics of its dividend policy.

101.5 Relationship Between the Optimal Payout Ratio and the Growth Rate

In this section, we investigate the relationship between the optimal payout ratio and the growth rate in terms of both exact and approximate approaches. Taking the partial derivative of Eq. (26) with respect to the growth rate, we obtain

$$\begin{aligned} \frac{\partial [\bar{D}(t)/\bar{x}(t)]}{\partial h} &= \left(-\frac{1}{a + b\bar{l}} \right) \left[\frac{-k + he^{(h-k)(T-t)}}{h - k} \right] + \left(1 - \frac{h}{a + b\bar{l}} \right) \\ &\quad \times \left[\frac{[(-k) + h(h - k)(T - t)]e^{(h-k)(T-t)} + k}{(h - k)^2} \right]. \end{aligned} \quad (27)$$

⁹See Hamilton (1994), pp. 45–46.

The sign of Eq. (27) is not only affected by the growth rate (h) but also by the expected rate of return on assets ($a + b\bar{l}$), the duration of future dividend payments ($T - t$), and the cost of capital (k).

Since the sign of Eq. (27) cannot be analytically determined, we use a sensitivity analysis approach to investigate the sign of Eq. (27). Table 1 shows the sign of partial derivatives of Eq. (27) under different values of the growth rate and the rate of return on assets as well as duration ($T - t$). We find that the relationship between the optimal payout ratio and the growth rate is always negative when the growth rate is higher than the rate of return on assets. If the growth rate is lower than the rate of return on asset, the direction of relationship essentially depends on the duration of the dividend payment ($T - t$). We find that the sign of Eq. (27) is negative if the duration ($T - t$) is small and the growth rate and rate of return on assets are within a reasonable range. In addition, the curved lines in Table 1 also indicate a nonlinear relationship between the growth rate and the optimal payout ratio. We can therefore conclude that the relationship between the optimal payout ratio and the growth rate is nonlinear and generally negative.

Table 1 shows the sensitivity analysis of the relationship between the optimal payout and the growth rate. Panel A presents, when the duration of dividend payments ($T - t$) is equal to 3 years and the cost of capital is equal to 7%, the values of Eq. (32) under different settings of the rate of return on assets and the growth rate. Similar to Panel A, Panel B and Panel C present the values of Eq. (32) when the duration of dividend payments is equal to 3 years and 5 years and the cost of capital is equal to 7%.

Based upon Eq. (27), Fig. 1 plots the change in the optimal payout ratio with respect to the growth rate in different durations of dividend payments and costs of capital. We find a negative relationship between the optimal payout ratio and the growth rate, indicating that a firm with a higher rate of return on assets tends to pay out less when its growth opportunities increase. Moreover, a firm with a lower growth rate and higher expected rate of return will not decrease its payout when its growth opportunities increase, but a firm with a lower growth and a higher expected rate of return on asset is not a general case in the real world. We also find that the duration of future dividend payments is an important determinant of the dividend payout decision, but the effect on the cost of capital is relatively minor.

In the finite growth case, if $(h - k)(T - t) < 1$, then following the MacLaurin expansion, the optimal payout ratio under no change in risk defined in Eq. (26) can be written as

$$[\bar{D}(t)/\bar{x}(t)] \approx \left(1 - \frac{h}{a + b\bar{l}}\right)(1 + h(T - t)) \quad (28)$$

The partial derivative of Eq. (28) with respect to the growth rate is

$$\frac{\partial [\bar{D}(t)/\bar{x}(t)]}{\partial h} \approx \left(\frac{[(a + b\bar{l}) - h](T - t) - h(T - t) - 1}{a + b\bar{l}} \right) \quad (29)$$

Table 1 Sensitivity analysis of the relationship between the optimal payout and the growth rate

Panel A. $T-t = 3$, Cost of Capital = 7%								
Growth (%)	ROA = 5%	ROA = 10%	ROA = 15%	ROA = 20%	ROA = 25%	ROA = 30%	ROA = 35%	ROA = 40%
1	-18.32	-7.77	-4.25	-2.49	-1.44	-0.73	-0.23	0.15
2	-19.39	-8.26	-4.55	-2.70	-1.58	-0.84	-0.31	0.08
3	-20.52	-8.78	-4.87	-2.92	-1.74	-0.96	-0.40	0.02
4	-21.69	-9.32	-5.20	-3.14	-1.91	-1.08	-0.49	-0.05
5	-22.91	-9.89	-5.55	-3.38	-2.08	-1.21	-0.59	-0.13
6	-24.19	-10.49	-5.92	-3.63	-2.26	-1.35	-0.70	-0.21
7	-25.56	-11.12	-6.31	-3.90	-2.46	-1.50	-0.81	-0.29
8	-26.92	-11.75	-6.70	-4.17	-2.65	-1.64	-0.92	-0.38
9	-28.38	-12.43	-7.12	-4.46	-2.86	-1.80	-1.04	-0.47
10	-29.90	-13.14	-7.55	-4.76	-3.09	-1.97	-1.17	-0.57
11	-31.48	-13.88	-8.01	-5.08	-3.32	-2.14	-1.31	-0.68
12	-33.14	-14.65	-8.49	-5.41	-3.56	-2.33	-1.45	-0.79
13	-34.86	-15.46	-8.99	-5.76	-3.81	-2.52	-1.60	-0.90
14	-36.66	-16.30	-9.51	-6.12	-4.08	-2.72	-1.75	-1.03
15	-38.54	-17.18	-10.06	-6.50	-4.36	-2.94	-1.92	-1.16
16	-40.50	-18.10	-10.63	-6.89	-4.65	-3.16	-2.09	-1.29
17	-42.54	-19.05	-11.22	-7.31	-4.96	-3.39	-2.28	-1.44
18	-44.67	-20.05	-11.85	-7.74	-5.28	-3.64	-2.47	-1.59
19	-46.89	-21.09	-12.49	-8.20	-5.62	-3.90	-2.67	-1.75
20	-49.20	-22.18	-13.17	-8.67	-5.97	-4.17	-2.88	-1.91
Panel B of $T-t = 5$, Cost of Capital = 7%								
Growth (%)	ROA = 5%	ROA = 10%	ROA = 15%	ROA = 20%	ROA = 25%	ROA = 30%	ROA = 35%	ROA = 40%
1	-17.33	-6.45	-2.83	-1.01	0.07	0.80	1.32	1.70
2	-18.99	-7.18	-3.24	-1.27	-0.09	0.70	1.26	1.68
3	-20.77	-7.96	-3.68	-1.55	-0.27	0.59	1.20	1.66

4	-22.70	-8.80	-4.17	-1.85	-0.46	0.46	1.13	1.62
5	-24.76	-9.71	-4.69	-2.18	-0.68	0.33	1.04	1.58
6	-26.97	-10.69	-5.26	-2.54	-0.91	0.17	0.95	1.53
7	-29.44	-11.78	-5.89	-2.95	-1.18	< 0.01	0.84	1.47
8	-31.90	-12.87	-6.53	-3.35	-1.45	-0.18	0.72	1.40
9	-34.63	-14.09	-7.24	-3.81	-1.76	-0.39	0.59	1.32
10	-37.57	-15.39	-8.00	-4.31	-2.09	-0.61	0.44	1.23
11	-40.71	-16.80	-8.83	-4.85	-2.45	-0.86	0.28	1.13
12	-44.07	-18.31	-9.72	-5.43	-2.85	-1.13	0.09	1.01
13	-47.67	-19.93	-10.68	-6.05	-3.28	-1.43	-0.11	0.88
14	-51.53	-21.66	-11.71	-6.73	-3.74	-1.75	-0.33	0.74
15	-55.66	-23.52	-12.81	-7.46	-4.25	-2.10	-0.57	0.57
16	-60.07	-25.52	-14.00	-8.25	-4.79	-2.49	-0.84	0.39
17	-64.79	-27.66	-15.28	-9.09	-5.38	-2.90	-1.14	0.19
18	-69.84	-29.95	-16.65	-10.01	-6.02	-3.36	-1.46	-0.03
19	-75.23	-32.40	-18.13	-10.99	-6.70	-3.85	-1.81	-0.28
20	-81.00	-35.03	-19.70	-12.04	-7.45	-4.38	-2.19	-0.55
Panel C of T-I = 15, Cost of Capital = 7%								
Growth (%)	ROA = 5%	ROA = 10%	ROA = 15%	ROA = 20%	ROA = 25%	ROA = 30%	ROA = 35%	ROA = 40%
1	-13.56	-1.52	2.49	4.50	5.71	6.51	7.08	7.51
2	-17.06	-2.56	2.27	4.69	6.14	7.11	7.80	8.31
3	-21.34	-3.89	1.93	4.84	6.58	7.75	8.58	9.20
4	-26.58	-5.57	1.43	4.93	7.03	8.43	9.43	10.18
5	-32.96	-7.69	0.73	4.94	7.47	9.15	10.36	11.26
6	-40.72	-10.34	-0.21	4.85	7.89	9.91	11.36	12.45
7	-51.15	-14.03	-1.66	4.53	8.24	10.72	12.48	13.81
8	-61.58	-17.72	-3.10	4.21	8.59	11.52	13.61	15.17
9	-75.40	-22.75	-5.21	3.57	8.83	12.34	14.85	16.73

(continued)

Table 1 (continued)

10	-92.10	-28.94	-7.89	2.63	8.95	13.16	16.17	18.42
11	-112.25	-36.53	-11.29	1.33	8.91	13.95	17.56	20.26
12	-136.53	-45.80	-15.55	-0.43	8.64	14.69	19.01	22.25
13	-165.74	-57.09	-20.88	-2.77	8.10	15.34	20.51	24.39
14	-200.85	-70.83	-27.49	-5.82	7.18	15.85	22.04	26.69
15	-243.01	-87.50	-35.67	-9.75	5.80	16.17	23.57	29.13
16	-293.57	-107.70	-45.74	-14.77	3.82	16.21	25.06	31.70
17	-354.16	-132.13	-58.11	-21.11	1.10	15.90	26.47	34.40
18	-426.70	-161.62	-73.25	-29.07	-2.57	15.11	27.73	37.20
19	-513.48	-197.17	-91.74	-39.02	-7.39	13.70	28.76	40.06
20	-617.19	-239.98	-114.24	-51.37	-13.65	11.49	29.46	42.93

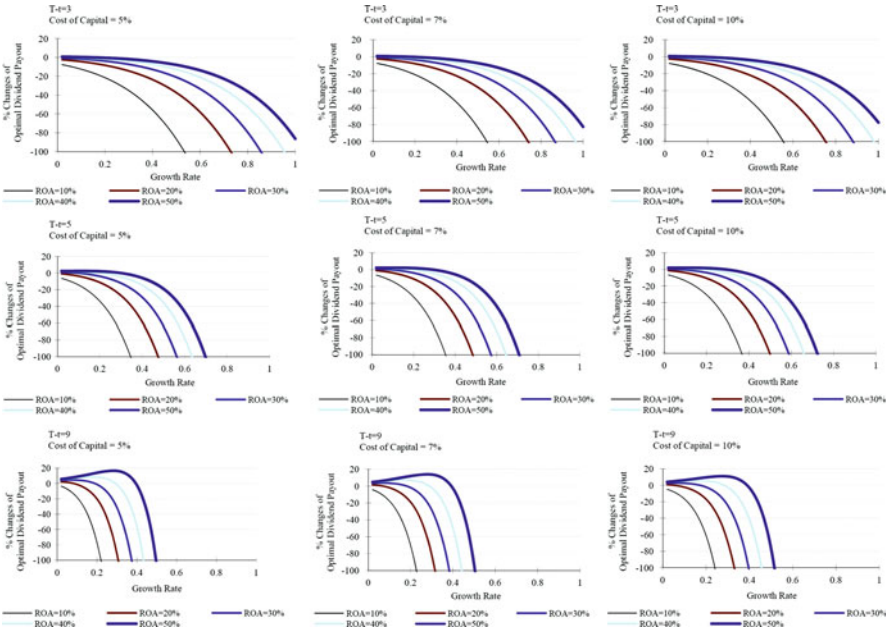


Fig. 1 Sensitivity analysis of the relationship between the optimal payout and the growth rate. The figures show the sensitivity analysis of the relationship between the optimal payout and the growth rate. In each figure, each line shows the percentage change of the optimal dividend payout with a 1% change in the growth rate. The various lines represent different levels of the rate of return on assets. The figures present different durations of dividend payments (T - t) and costs of capital

Equation (29) indicates that the relationship between the optimal dividend payout and the growth rate depends on firm’s level of growth, the rate of return on assets, and the duration of future dividend payment.¹⁰ Eq. (29) is negative when the rate of return on assets is lower than the growth rate. This implies that the firm will reduce its payout when its growth rate increases. The higher growth rate (h) and the lower rate of return on assets ($a + b\bar{l}$) will lead to a more negative relationship between dividend payout ratio and growth rate. More specifically, the condition of Eq. (29) will lead to a negative relationship between the optimal payout ratio and the growth rate.

$$h > \frac{1}{2} \left[(a + b\bar{l}) - \frac{1}{(T - t)} \right] \tag{30}$$

Consistent with the sensitivity analysis of Eq. (27), when a firm with a high growth rate or a low rate of return on assets faces a growth opportunity, it will decrease its dividend payout to generate more cash to meet such a new investment.

¹⁰Please see the Appendix C for the derivation of Eq. (28) and Eq. (29).

A possible explanation is that high growth firms need more retained earnings to meet their future growth opportunities because the growth rate is the main determinant of value in the case of such companies, but low growth firms do not need more earnings to maintain their low growth perspective and can afford to increase their payouts. Based on the flexibility concerns, the relationship between firms' payout ratios and their growth rates is therefore negative.

101.6 Relationship Between Optimal Payout Ratio and Risks

Equation (26) implies that the optimal payout ratio is a function of the expected profitability rate ($a + b\bar{l}$), growth rate (h), cost of capital (k), age ($T-t$), total risk ($\sigma(t)^2$), and the correlation coefficient between profit and market rate of return ($\rho(t)^2$). In addition, Eq. (26) is also a function of two dynamic variables – the relative time rate of change in the total risk of the firm, $[\dot{\sigma}(t)^2/\sigma(t)^2]$, and the relative time rate of change in the covariability of the firm's earnings with the market, $[\dot{\rho}(t)^2/\rho(t)^2]$. This theoretical dynamic relationship between the optimal payout ratio and other determinants can be used to do empirical studies to determine dividend policy. The dynamic effects of variations in $[\dot{\sigma}(t)^2/\sigma(t)^2]$ and $[\dot{\rho}(t)^2/\rho(t)^2]$ on the time path of optimal payout ratio can be investigated under the following four cases: (1) changes in total risk, (2) changes in correlation between profit and the market rate of return (i.e., systematic risk), (3) changes in total risk and systematic risk, and (4) no changes in risk.

101.6.1 Case 1: Total Risk

First, we examine the effect of $[\dot{\sigma}(t)^2/\sigma(t)^2]$ on the optimal payout ratio. By differentiating Eq. (26) with respect to $[\dot{\sigma}(t)^2/\sigma(t)^2]$, we obtain

$$\frac{\partial [\bar{D}(t)/\bar{x}(t)]}{\partial [\dot{\sigma}(t)^2/\sigma(t)^2]} = \left(1 - \frac{h}{a + b\bar{l}}\right) \left[\frac{e^{(h-k)(T-t)} - 1}{h - k}\right]. \quad (31)$$

In Eq. (31) the cost of capital, k , can be either larger or smaller than the growth rate h . We can show that $\frac{e^{(h-k)(T-t)} - 1}{h - k}$ is always larger than 0, regardless of whether k is larger or smaller than h .¹¹ Thus, the sign of Eq. (31) depends on the sign of $\left(1 - \frac{h}{a + b\bar{l}}\right)$, which depends on the growth rate h relative to $(a + b\bar{l})$.

¹¹If $h > k$, then $e^{(h-k)(T-t)} > 1$, and both the numerator and denominator are greater than zero, resulting in a positive value; if $h < k$, then $e^{(h-k)(T-t)} < 1$, and both the numerator and denominator are less than zero, resulting in a positive value. Thus, $\frac{e^{(h-k)(T-t)} - 1}{h - k}$ is always larger than 0, regardless of whether k is larger or smaller than h .

If the growth rate h is equal to $(a + b\bar{l})$, then $\left(1 - \frac{h}{a+b\bar{l}}\right)$ is equal to zero. Eq. (31) is thus zero, and the change in total risk will not affect the payout ratio because the first derivative of the optimized payout ratio, Eq. (26), with respect to $\left[\dot{\sigma}(t)^2/\sigma(t)^2\right]$ is always zero.

If growth rate h is larger than $(a + b\bar{l})$, then the entire first derivative of Eq. (26) with respect to $\left[\dot{\sigma}(t)^2/\sigma(t)^2\right]$ is negative (i.e., Eq. (31) is negative); furthermore, $h > (a + b\bar{l})$ implies that the growth rate of a firm is larger than its expected profitability rate. An alternative case is $h < (a + b\bar{l})$, which implies that the growth rate of a firm is less than its expected profitability rate. This situation can occur when a company is either in a low growth, no growth, or negative growth stage. Under this situation, a company will increase its payout ratio as shown in Eq. (31). If $h < (a + b\bar{l})$, then Eq. (31) is positive, indicating that a relative increase in the risk of the firm would increase its optimal payout ratio. This implies that a relative increase in the total risk of the firm would decrease its optimal payout ratio. Lintner (1965) and Blau and Fuller (2008) have found this kind of relationship, yet they did not theoretically show how it can be derived.

Jagannathan et al. (2000) empirically show that operation risk is negatively related to the propensity to increase payouts in general and dividends in particular. Our theoretical analysis in terms of Eq. (31) shows that the change in total risk is negatively or positively related to the payout ratio, conditional on the higher growth rate relative to the expected profitability rate. We find negative relationships between payout and the change in total risk for high growth firms ($h > (a + b\bar{l})$). A possible explanation is that in the case of high growth firms, a firm must reduce the payout ratio and retain more earnings to build up “precautionary reserves,” which become all the more important for a firm with volatile earnings over time. High growth firms thus tend to retain more earnings when they face higher risk. By contrast, in the case of established low growth firms ($h < (a + b\bar{l})$), low growth firms are likely to be more mature and have most likely already built such reserves over time. They probably do not need more earnings to maintain their low growth perspective and can afford to increase the payout; consequently, when facing higher risk on their earnings, low growth firms can reduce their shareholders’ risk by paying more dividends to their shareholders.

The age of the firm ($T-t$), which is one of the variables in Eq. (31), becomes an important factor because the very high growth firms are also the newer firms with very little built-up precautionary reserves.

Under more dynamic conditions, we provide further evidence of the validity of Lintner’s (1965) observations that, *ceteris paribus*, optimal dividend payout ratios vary directly with the variance of the firm’s profitability rates. The rationale for such relationships, even when the systematic risk concept is incorporated into the analysis, is obvious; that is, holding $\rho(t)^2$ constant and letting the $\sigma(t)^2$ increase imply that the covariance of the firm’s earnings with the market does not change though its relative proportion to the total risk increases.

101.6.2 Case 2: Systematic Risk

To examine the effect of a relative change in $\left[\dot{\rho}(t)^2/\rho(t)^2\right]$ (i.e., systematic risk) on the dynamic behavior of the optimal payout ratio, we differentiate Eq. (26) to obtain

$$\frac{\partial [\bar{D}(t)/\bar{x}(t)]}{\partial [\dot{\rho}^2(t)/\rho^2(t)]} = \left(1 - \frac{h}{a + b\bar{l}}\right) \left[\frac{e^{(h-k)(T-t)} - 1}{h - k}\right]. \quad (32)$$

The sign of Eq. (32) can be analyzed as with Eq. (31), so the conclusions of Eq. (32) are similar to those of Eq. (31). A relative change in $\rho(t)^2$ can either decrease or increase the optimal payout ratio, all things being equal. The effect of non-stationarity in the firm's nondiversifiable risk would tend to be obliterated should both the systematic and the unsystematic components of total risk not be clearly identified in the expression for optimal payout ratio. Although the total risk of the firm is stationary (i.e., $[\dot{\sigma}(t)^2/\sigma(t)^2]$ is equal to zero) a change in the total risk complexion of the firm could still conceivably, occur because of an increase or decrease in the covariability of its earnings with the market. Eq. (26) and Eq. (32) clearly identify the effect of such a change in the risk complexion of the firm on its optimal payout ratio.

An examination of Eq. (26) indicates that only when the firm's earnings are perfectly correlated with the market (i.e., $\rho^2 = 1$), whether the management arrives at its optimal payout ratio using the total variance concept of risk or the market concept of risk does not matter. For every other case, the optimal payout ratio followed by management using the total variance concept of risk would be an overestimate of the true optimal payout ratio for the firm based on the market concept of risk underlying the capital asset pricing theory.

Management may decide not to use the truly dynamic model and instead substitute an average of the long run systematic risk of the firm, but for $\dot{\rho}^2(t) > 0$, because the initial average is higher than the true $\rho^2(t)$, the management would pay out less or more in the form of dividends than is optimal. In other words, the payout ratio followed in the initial part of the planning horizon would be an overestimate or an underestimate of the optimal payout under truly dynamic specifications.

Rozeff (1982) empirically shows a negative relationship between the β coefficient (systematic risk) and the payout level. The theoretical analysis in terms of Eq. (32) provides a more detailed analytical interpretation of his findings. The explanations of these results are similar to those discussed for findings in previous sections.

101.6.3 Case 3: Total Risk and Systematic Risk

In our third case, we attempt to investigate the compounded effect of a simultaneous change in the total risk of the firm and also a change in its decomposition into the

market-dependent and market-independent components. Taking the total differential of Eq. (26) with respect to $\left[\dot{\sigma}(t)^2/\sigma(t)^2\right]$ and $\left[\dot{\rho}(t)^2/\rho(t)^2\right]$, we obtain

$$d[\overline{D}(t)/\overline{x}(t)] = \gamma d\left(\frac{\dot{\sigma}(t)^2}{\sigma(t)^2}\right) + \gamma d\left(\frac{\dot{\rho}(t)^2}{\rho(t)^2}\right), \quad (33)$$

where $\gamma = \left(1 - \frac{h}{a+b\bar{l}}\right) \left[\frac{e^{(h-k)(T-t)} - 1}{h-k}\right]$. Also, γ can be either negative or positive, as shown above.

Now from Eq. (33), the greatest decrease or increase in the optimal payout ratio would obviously occur when both $\dot{\sigma}(t)^2$ and $\dot{\rho}(t)^2$ are positive. This implies that the total risk of the firm increases, and in addition, its relative decomposition into systematic and unsystematic components also changes, making the firm's earnings still more correlated with the market. Under this circumstance, the decrease or increase in the optimal payout would now represent the compounded effect of both these changes; however, it is conceivable that although $\dot{\sigma}(t)^2$ is positive, $\dot{\rho}(t)^2$ is negative, tending to offset the decrease or increase in the optimal payout ratio resulting from the former. Alternatively, $\dot{\sigma}(t)^2$ could be negative, indicating a reduction in the total risk of the firm and may offset the increase in the optimal payout ratio resulting from a positive $\dot{\rho}(t)^2$.

To what extent the inverse variations in the total risk and the risk complexion of the firm will offset each other's effects on the optimal payout ratio for the firm would, of course, be dependent upon the relative magnitudes of $\dot{\rho}(t)^2$ and $\dot{\sigma}(t)^2$. To see the precise trade-off between the two dynamic effects of $\left[\dot{\sigma}(t)^2/\sigma(t)^2\right]$ and $\left[\dot{\rho}(t)^2/\rho(t)^2\right]$ on the optimal payout ratio, let the total differential of (26), given in Eq. (33), be set equal to zero, yielding

$$d\left[\dot{\sigma}(t)^2/\sigma(t)^2\right] = -d\left[\dot{\rho}(t)^2/\rho(t)^2\right] \quad (34)$$

Equation (34) implies that the relative increase (or decrease) in $\sigma(t)^2$ has a one-to-one correspondence with the relative decrease (or increase) in $\rho(t)^2$, so in Eq. (34) conditions are established for relative changes in $\rho(t)^2$ and $\sigma(t)^2$, which lead to a null effect on the optimal dividend payout ratio.

101.6.4 Case 4: No Change in Risk

Now we consider the least dynamic situation, in which no changes exist in total risk or systematic risk, assuming $\dot{\sigma}(t)^2 = 0$ and $\dot{\rho}(t)^2 = 0$. Under this circumstance, Eq. (26) reduces to

$$[\overline{D}(t)/\overline{x}(t)] = \left(1 - \frac{h}{a+b\bar{l}}\right) \left[\frac{-k + he^{(h-k)(T-t)}}{h-k}\right]. \quad (35)$$

Thus, when the firm's total risk and covariability of its earnings with the market are assumed stationary, Eq. (35) indicates that a firm's optimal payout ratio is independent of its risk. Notice that neither $\sigma(t)^2$ nor $\rho(t)^2$ now appear in the expression for the optimal payout ratio given in Eq. (35). These conclusions, like those of Wallingford (1972a, b) for example, run counter to the intuitively appealing and well-accepted theory of finance emphasizing the relevance of risk for financial decision-making.¹² Our model clearly shows that the explanation for such unacceptable implications of the firm's total risk and its market-dependent and market-independent components for the firm's optimal payout policy lies, of course, in the totally unrealistic assumptions of stationarity underlying the derivation of such results as illustrated in Eq. (35).

101.7 Empirical Evidence

A growing body of literature focuses on the determinants of optimal dividend payout policy. Rozeff (1982), Jagannathan et al. (2000), Grullon et al. (2002), Aivazian et al. (2003), Blau and Fuller (2008), and others empirically investigate the determination of dividend policy, but none of them has a solid theoretical model to support their findings. Based upon our theoretical model and its implications discussed in the foregoing sections, we develop the three testable hypotheses that follow.

Hypothesis 1: Firms generally reduce their dividend payouts when their growth rates increase.

The negative relationship between the payout ratio and the growth ratio in our theoretical model implies that high growth firms need to reduce the payout ratio and retain more earnings to build up "precautionary reserves," but low growth firms are likely to be more mature and already build up their reserves for flexibility considerations. Rozeff (1982), Fama and French (2001), Blau and Fuller (2008), and others argue that high growth firms will have higher investment opportunities and tend to pay out less in dividends. Based upon flexibility concerns, we predict that high growth firms pay higher dividends. This result is obtained when risk factor is not explicitly considered.

Following Eqs. (31) and (32), we theoretically find that the relationship between the payout ratio and the risk can be either negative or positive, depending upon whether the growth rate is higher or lower than the rate of return on total assets. Based upon this finding, we develop two other hypotheses.

Hypothesis 2: The relationship between the firms' dividend payouts and their risks is negative when their growth rates are higher than their rates of return on asset.

High growth firms need to reduce the payout ratio and retain more earnings to build up "precautionary reserves," which become more important for a firm with volatile earnings over time. For flexibility considerations, high growth firms tend to

¹²For example, Lintner (1963).

retain more earnings when they face higher risk. This theoretical result is consistent with the flexibility hypothesis.

Hypothesis 3: The relationship between the firms' dividend payouts and their risks is positive when their growth rates are lower than their rates of return on asset.

Low growth firms are likely to be more mature and have most likely already built such reserves over time, and they probably do not need more earnings to maintain their low growth perspective and can afford to increase the payout (see Grullon et al. 2002). Because the higher risk may involve higher cost of capital and make the free cash flow problem worse, for free cash flow considerations, low growth firms tend to pay more dividends when they face higher risk. This theoretical result is consistent with the free cash flow hypothesis.

101.7.1 Sample Description

We collect the firm information, including total asset, sales, net income, and dividends payout, from Compustat. Stock price, stock returns, share codes, and exchange codes are retrieved from the Center for Research in Security Prices (CRSP) files. The sample period is from 1969 to 2009. Only common stocks (SHRCD = 10, 11) and firms listed on NYSE, AMEX, or NASDAQ (EXCE = 1, 2, 3, 31, 32, 33) are included in our sample. We exclude utility services (SICH = 4900–4999) and financial institutions (SICH = 6000–6999).¹³ The sample includes those firm-years with at least 5 years of data available to compute average payout ratios, growth rate, return on assets, beta, total risk, size, and book-to-market ratios. The payout ratio is measured as the ratio of the dividend payout to the net income. The growth rate is the sustainable growth rate proposed by Higgins (1977). The beta coefficient and total risk are estimated by the market model over the previous 60 months. For the purpose of estimating their betas, firm-years in our sample should have at least 60 consecutive previous monthly returns. To examine the optimal payout policy, only firm-years with five consecutive dividend payouts are included in our sample.¹⁴ Considering the fact that firm-years with no dividend

¹³We filter out those financial institutions and utility firms based on the historical Standard Industrial Code (SIC) available from COMPUSTAT. When a firm's historical SIC is unavailable for a particular year, the next available historical SIC is applied instead. When a firm's historical SIC is unavailable for a particular year and all the years after, we use the current SIC from COMPUSTAT as a substitute.

¹⁴To avoid creating a large difference in dividend policy, on one hand managers partially adjust firms' payout by several years to reduce the sudden impacts of the changes in dividend policy. On the other hand, they use not only 1-year firm conditions but also multiyear firm conditions to decide how much they will pay out. In examining the optimal payout policy, we use the 5-year rolling averages for all variables.

payout 1 year before (or after) might not start (or stop) their dividend payouts in the first (fourth) quarter of the year, we exclude from our sample firm-years with no dividend payouts 1 year before or after to ensure the dividend payout policy reflects the firm's full-year condition.

Table 2 shows the summary statistics for 2,645 sample firms during the period from 1969 to 2009. Panel A of Table 2 lists the number of firm-year observations for all sample high growth firms and low growth firms, respectively. High growth firm-years are those firm-years that have 5-year average sustainable growth rates higher than their 5-year average rate of return on assets. Low growth firm-years are those firms with 5-year average sustainable growth lower than their 5-year average rate of return on assets. The sample size increases from 345 firms in 1969 to 1,203 firms in 1982, while declining to 610 firms by 2009. A total of 28,333 dividend paying firm-years are included in the sample. When classifying high growth firms and low growth firms relative to their return on assets, the proportion of high growth firms increases over time. The proportion of firm-years with a growth rate higher than return on assets increases from less than 50% during the late 1960s and early 1970s to 80% in 2008. Panel B of Table 2 shows the 5-year moving averages of mean, median, and standard deviation values for the measures of payout ratio, growth rate, rate of return, beta coefficient, total risk, market capitalization, and market-to-book ratio across all firm-years in the sample. Among high growth firms, the average growth rate is 12.33%, and the average payout ratio is 31.80%; but for low growth firms, the average growth rate is 4.13%, and the average payout ratio is 57.62%. High growth firms undertake more beta risk and total risk, indicating that high growth firms undertake both more systematic risk and unsystematic risk to pursue a higher rate of return.

Table 2 presents the descriptive statistics for those major characteristics of our sample firms. Sample includes those firms listed on NYSE, AMEX, and NASDAQ with at least 5 years of data available to compute average payout ratios, growth rate, return on assets, beta, total risk, size, and book-to-market ratios. All financial service operations and utility companies are excluded. Panel A lists the numbers of firm-years observations for all sample firms, high growth firms, and low growth firms, respectively, during the period between year 1969 and year 2009. High growth firm-years are defined as firm-years with sustainable growth rates higher than their rates of return on assets. Low growth firm-years are defined as firm-years with sustainable growth rates lower than their rate of return on assets. Panel B lists the mean, median, and standard deviation values of the 5-year average of the payout ratio, growth rate, rate of return on assets, beta risk, total risk, size, and book-to-market ratio. The payout ratio is measured as the ratio of the dividend payout to the earnings. Growth rate is the sustainable growth rate proposed by Higgins (1977). The beta coefficient and total risk are estimated by the market model over the previous 60 months. Size is defined as market capitalization calculated by the closing price of the last trading day of June of that year times the outstanding shares at the end of June of that year.

Table 2 Summary statistics of sample firm characteristics

Panel A. Sample size						
Year	Number of firm years		Year		Number of firm years	
	All	Growth > ROA	Growth < ROA		All	Growth > ROA
1969	345	161	184	1990	690	522
1970	360	175	185	1991	668	511
1971	404	201	203	1992	653	494
1972	513	269	244	1993	642	460
1973	535	308	227	1994	655	479
1974	572	371	201	1995	651	483
1975	609	432	177	1996	693	530
1976	650	486	164	1997	725	582
1977	678	530	148	1998	743	620
1978	711	553	158	1999	725	612
1979	779	620	159	2000	709	607
1980	764	636	128	2001	659	569
1981	929	785	144	2002	599	503
1982	1,203	1,003	200	2003	571	475
1983	1,151	933	218	2004	525	433
1984	1,067	832	235	2005	481	391
1985	1,010	744	266	2006	510	430
1986	958	669	289	2007	542	451
1987	897	645	252	2008	579	470
1988	847	615	232	2009	610	484
1989	721	531	190	All years	28,333	21,065

(continued)

Table 2 (continued)

Panel B of Descriptive statistics of characteristics of sample							
All sample (N = 28,333)							
	Payout ratio	Growth rate	ROA	Beta	Total risk	Size (\$MM)	M/B
Mean	0.3793	0.1039	0.0723	1.0301	0.0106	3,072	1.7940
Median	0.3540	0.0886	0.0648	1.0251	0.0089	291	1.3539
Stdev	0.1995	0.7444	0.0389	0.4272	0.0078	14,855	1.9479
High growth firms (N = 21,065)							
	Payout ratio	Growth rate	ROA	Beta	Total risk	Size (\$MM)	M/B
Mean	0.3180	0.1233	0.0698	1.0624	0.0112	3,267	1.7951
Median	0.2996	0.1002	0.0638	1.0581	0.0095	314	1.3757
Stdev	0.1658	0.8060	0.0355	0.4352	0.0070	15,806	1.6496
Low growth firms (N = 6,728)							
	Payout ratio	Growth rate	ROA	Beta	Total risk	Size (\$MM)	M/B
Mean	0.5762	0.0413	0.0800	0.9265	0.0087	2,447	1.7904
Median	0.5542	0.0524	0.0692	0.9375	0.0071	229	1.3007
Stdev	0.1690	0.4918	0.0476	0.3822	0.0099	11,250	2.6909

Table 3 Payout ratios partitioned by growth rate and risks

Panel A. Payout ratios partitioned by growth rate						
		<i>Low growth</i>				<i>High growth</i>
		1	2	3	4	5
<i>All sample</i>	<i>Mean</i>	0.4864	0.3835	0.3272	0.2696	0.2213
	<i>Median</i>	0.4835	0.3795	0.3196	0.2556	0.1967
Panel B. Payout ratios partitioned by beta risk						
		<i>Low beta</i>				<i>High beta</i>
		1	2	3	4	5
<i>All sample</i>	<i>Mean</i>	0.3912	0.3940	0.3963	0.3609	0.3371
	<i>Median</i>	0.3716	0.3765	0.3737	0.3328	0.3009
<i>High growth firms</i>	<i>Mean</i>	0.3173	0.3229	0.3204	0.3025	0.2886
	<i>Median</i>	0.3027	0.3081	0.3028	0.2874	0.2588
<i>Low growth firms</i>	<i>Mean</i>	0.6115	0.6064	0.6182	0.6189	0.6292
	<i>Median</i>	0.5864	0.5748	0.5947	0.6042	0.6090
Panel C. Payout ratios partitioned by total risk						
		<i>Low total risk</i>				<i>High total risk</i>
		1	2	3	4	5
<i>All sample</i>	<i>Mean</i>	0.4274	0.4058	0.3618	0.3285	0.3041
	<i>Median</i>	0.4099	0.3808	0.3333	0.2899	0.2588
<i>High growth firms</i>	<i>Mean</i>	0.3499	0.3286	0.2957	0.2698	0.2531
	<i>Median</i>	0.3381	0.3138	0.2745	0.2431	0.2190
<i>Low growth firms</i>	<i>Mean</i>	0.6078	0.6202	0.6269	0.6091	0.6115
	<i>Median</i>	0.5732	0.5942	0.6062	0.5863	0.6025

101.8 Univariate Analysis

To examine Hypothesis 1, we divide our sample into five groups by growth rate. As Table 3 Panel A indicates, the average (median) payout ratio is 48.64% (48.35%) for the lowest growth group and is 22.13% (19.67%) for the highest growth group. The argument of “precautionary reserves” for flexibility concerns and Hypothesis 1 can therefore be confirmed. We further divide our sample into 5 groups by beta risk and total risk indicated in Panel B and Panel C of Table 3. We can observe a monotonic decrease in payout ratios when the beta risk or the total risk increases, which is consistent with the findings of Rozeff (1982), Fenn and Liang (2001), Grullon et al. (2002), Aivazian et al. (2003), and Blau and Fuller (2008), but they do not consider the relationship between the payout ratio and the risk may alter by the growth. To examine the Hypothesis 2 and Hypothesis 3, we further divide our sample by two-way sorts on the growth rate and risks. High growth groups are firm-years with sustainable growth rates higher than their rate of return on total assets. Low growth groups are firm-years with sustainable growth rates lower than their rate of return on total assets. In Table 3 Panel B, among high growth firms, the average (median) payout ratio decreases from 31.73% (30.10%) to 28.86% (25.88%) when

the beta risk increases. Among low growth firms, however, the average (median) payout ratio increases from 61.15% (58.64%) to 62.92% (60.90%) when the beta risk increases. Similar to Panel B, Panel C also shows a negative relationship between the payout ratio and the total risk among high growth firms and a positive relationship between the payout ratio and the total risk among low growth firms. Above all, the static analysis results of Panel B and Panel C support Hypotheses 2 and 3 that the relationship between the payout ratio and the risk depends upon the growth rate of a firm.

Table 3 presents the average and the median payout ratios in difference groups partitioned by growth rate, beta risk, or total risk during the sample period from 1969 to 2009. Panel A reports the average and the median payout ratios by one-way sort on the growth rate. Panel B reports the average and the median payout ratios by independent two-way sort on the growth rate and the beta. Panel C reports the average and the median payout ratios by independent two-way sort on the growth rate and the total risk. High growth firm-years are defined as firm-years with sustainable growth rates higher than their rates of return on assets. Low growth firm-years are defined as firm-years with sustainable growth rates lower than their rate of return on assets.

101.8.1 Multivariate Analysis

To examine the relationship between the payout ratio and other financial variables, we propose fixed-effects models of the payout ratio as follows¹⁵:

$$\ln \left(\frac{\text{payout ratio}_{i,t}}{1 - (\text{payout ratio}_{i,t})} \right) = \alpha + \beta_1 \text{Risk}_{i,t} + \beta_2 D_{i,t}(g_{i,t} < c \cdot ROA_{i,t}) \cdot \text{Risk}_{i,t} + \beta_3 \text{Growth}_{i,t} + \beta_4 \text{Risk}_{i,t} \times \text{Growth}_{i,t} + \beta_5 \ln(\text{Size})_{i,t} + \beta_6 ROA_{i,t} + e_{i,t}. \quad (36)$$

In the regression, the dependent variable is the logistic transformation of the payout ratio. Independent variables include risk measure (beta coefficient or total risk), the interaction of dummy variable and risk measure, growth rate, the interaction of risk measure and growth rate, log of size, and rate of return on total assets.¹⁶ Based upon the theoretical model and its implications from Sect. 101.6, we assume

¹⁵ The dummy variable $D_{i,t}(g_{i,t} < c \cdot ROA_{i,t})$ used in Eq. (36) implies that the relationship between the payout ratio and risks is nonlinear (piecewise regression). In other words, the breakpoint of the structural change is at $g_{i,t} = c \cdot ROA_{i,t}$. Based upon our theoretical model, we assume that c is equal to 1 in our empirical work.

¹⁶ Besides merely adding an interaction dummy as indicated in Eq. (36), we include an intercept dummy to take care of the individual effect of two groups. We also run regressions for high growth firms and low growth firms separately. Results from both models are qualitatively the same as those from Eq. (36) and also support Hypotheses 1 to 3.

that c is equal to 1. The dummy variable (D_i) is equal to 1 if a firm's 5-year average growth rate is less than its 5-year average rate of return on assets and 0 otherwise. Such structure allows us to analyze the relationship between payout ratio and growth rate and the relationship between payout ratio and risk under different growth rate levels.

Thompson (2011), Peterson (2009), and Cameron et al. (2006) have pointed out that standard errors of two-way fixed-effects estimates can be biased if 2-dimensional clustering (clustering in the cross-sectional errors and clustering in time-series errors) is not controlled for. Thompson (2011) and Boehmer et al. (2010) have empirically found that these clustering effects are not important for large samples.

The robust standard errors are very similar to standard errors for ordinary least square (OLS), suggesting that the fixed effects and control variables are removing most of the correlation that is present across observations. In addition, our dataset cannot be meaningfully applied to the clustering effect model¹⁷; therefore, statistical inferences in this chapter are conducted using fixed-effects standard errors.

Table 4 provides the results of fixed-effects regressions for 2,645 firms during the period 1969 to 2009. *Model (1)* and *Model (2)* show that the estimated coefficients of the growth rate are -0.03 with a t -statistics of -4.85 and -0.03 with a t -statistics of -4.85 , respectively. Such significantly negative coefficients confirm Hypothesis 1, which states that high growth firms will pay less in dividends for the consideration of flexibility. We also include an interaction term of risk and growth rate into *Model (3)* and *Model (4)*. The results in *Model (3)* and *Model (4)* also support Hypothesis 1.

Table 4 presents results from fixed-effects regressions of dividend payout ratios on firm characteristics for 2,645 firms during the period from 1969 to 2009. The regressions are as follows:

$$\ln \left(\frac{\text{payout ratio}_{i,t}}{1 - (\text{payout ratio}_{i,t})} \right) = \alpha + \beta_1 \text{Risk}_{i,t} + \beta_2 D(g_{i,t} < c \cdot ROA_{i,t}) \cdot \text{Risk}_i \\ + \beta_3 \text{Growth}_{i,t} + \beta_4 \text{Risk}_{i,t} \times \text{Growth}_{i,t} \\ + \beta_5 \ln(\text{Size})_{i,t} + \beta_6 ROA_{i,t} + e_{i,t}.$$

The dependent variable is the payout ratio with a logistic transformation. Following the theoretical model, we assume c is equal to 1. The dummy variable is equal to 1 if a firm's 5-year average growth rate is less than its 5-year average ROA, and 0 otherwise. The independent variables are beta risk (total risk), dummy times beta risk (total risk), interaction between growth rate and beta risk (total risk), growth rate, log of size, and the rate of return on assets. t -statistics are presented in

¹⁷Because our sample is an unbalanced panel data, the clustering computer program cannot meaningfully estimate the variance components, variance of firm ($\hat{V}_{firm}$), variance of time ($\hat{V}_{time}$), and heteroskedasticity-robust OLS variance ($\hat{V}_{white}$).

Table 4 Fixed-effects regressions using theoretical structural change point

Model	Intercept	Beta	D*beta	Growth* beta	Total risk	D*total risk	Growth* total risk	Growth	ln(Size)	ROA	Adj- R ²	F-test (p-value)
(1)	0.32*** (2.70)	-0.18*** (-9.82)						-0.03*** (-4.85)	0.04*** (3.76)	-13.24*** (-63.14)	0.1945	
(2)	0.45*** (3.76)				-17.47*** (-14.64)			-0.03*** (4.68)	0.02** (2.10)	-13.07*** (-62.42)	0.1982	
(3)	0.31*** (2.64)	-0.14*** (-7.76)		-0.29*** (-8.84)				0.09*** (6.38)	0.04*** (3.47)	-12.96*** (-61.17)	0.1970	
(4)	0.45*** (3.77)				-17.14*** (-13.13)		-2.90 (-0.63)	<0.01 (0.01)	0.02** (2.06)	-13.06*** (-61.97)	0.1982	
(5) c = 1	0.33*** (2.94)	-0.23*** (-13.88)	0.71*** (57.57)					-0.02*** (-3.70)	0.002 (0.21)	-11.54*** (-57.83)	0.2867	<0.01
(6) c = 1	0.52*** (4.58)				-27.81*** (-24.15)	63.35*** (52.23)		-0.02*** (-3.53)	-0.002 (-0.27)	-11.40*** (-56.55)	0.2753	<0.01
(7) c = 1	0.32*** (2.91)	-0.22*** (-12.69)	0.70*** (56.96)	-0.13*** (-4.02)				0.03*** (2.35)	-0.001 (-0.10)	-11.43*** (-56.76)	0.2871	<0.01
(8) c = 1	0.51*** (4.53)				-29.61*** (-23.44)	63.68*** (52.35)	15.30*** (3.46)	-0.16*** (-3.89)	-0.001 (-0.06)	-11.47*** (-56.63)	0.2756	<0.01

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively

parentheses, and the last column presents p -values of the F -test on the null hypothesis that the restricted *Models (1) to (4)* are not different from the restricted *-Models (5) to (8)*.

Models (1) to (4) show that the relationship between the payout ratio and the risk is significantly negative. The results are similar to the findings of Rozeff (1982), Jagannathan et al. (2000), and Grullon et al. (2002), indicating that dividend payouts are negatively correlated to firm risks; but our theoretical model shows that if firms follow their optimal dividend payout policy, the relationship between dividend payouts and firm risks depends on their growth rates relative to their rate of return on total assets as our theoretical analysis presented in Sect. 101.6. In Table 2, we find the number of firms with a higher growth rate with respect to their rate of return on assets is greater than the number of firms with a lower growth rate with respect to their rates of return on assets. When pooling high growth firms and low growth firms together, the negative risk effect of high growth firms will dominate the positive risk effect of low growth firms due to the larger proportion of high growth firms in the observations. The results of the negative relationship between the payout ratio and the risk shown in *Models (1) to (4)* may therefore result from the greater proportion of high growth firms. Based on our subsequent analysis, the effect of growth rates on dividend payout policies can be more accurately found when firms are separated into high growth firms and low growth firms relative to their rates of return on total assets of return on assets.

To test Hypotheses 2 and 3, we introduce an interaction term of the dummy variable and the risk. In *Model (5)* and *Model (6)*, the estimated coefficients of risk are -0.23 with a t -statistics of -13.88 and -0.22 with a t -statistics of -12.69 , respectively. The significantly negative coefficients support the hypothesis 2 that, because of the consideration of flexibility, the payout ratio and the risk are negatively correlated for firms with a higher growth rate relative to their rate of return on assets. In addition, significant and positive coefficients of the interaction term of the dummy variable and the risk indicate that, when the risk changes, the dividend policy for low growth firms is different from that of high growth firms. By summing the coefficient of risk and the coefficient of interaction term, we can obtain coefficients of 0.48 and 35.54 for beta and total risk, respectively, indicating the relationship between the payout ratio and the risk for low growth firms is positive. That is, when the risk increases, low growth firms will follow their optimal payout policies to increase their dividend payouts. Hypothesis 3 is thus confirmed in our empirical work. Blau and Fuller (2008) find that the flexibility hypothesis is more suitable than the free cash flow hypothesis to explain the dividend policy, but their method cannot separate the dividend policy decisions between high growth firms and low growth firms. Consequently, the model used in this chapter can be regarded as a generalization of their results.

In *Model (7)* and *Model (8)* in Table 4, we further include the interaction term of the risk and the growth rate into the regressions. The results still support our hypotheses that the relationship between the payout ratio (Hypothesis 1) and the growth rate is negative and the relationship between the payout ratio and the risk depends on firm's growth rate with respect to its rate of return on total assets

(Hypotheses 2 and 3). In addition, we find that the interaction terms of the risk and the growth rate are significantly different from zero. We also find that the adjusted R-squares for *Models (5) to (8)* are higher than those for models without dummies. *F* tests also reject the null hypothesis that the regression with the interaction of the dummy variable and the risk is not different from the regression without the interaction of the dummy and the risk. We can thus conclude that the payout ratio is not linearly related to the growth rate or to the risk, and previous empirical studies on dividend policy using linear model may suffer from model misspecification.

101.8.2 Moving Estimates Process for Structural Change Model

In Eqs. (31) and (32) in Sect. 101.6, we theoretically show that the structural change breakpoint for the relationship between the payout ratio and risks is at $g_{i,t} = ROA_{i,t}$. In empirical works, we use dummy variable approach to separate the sample into a high growth ($g_{i,t} > c \cdot ROA_{i,t}$) group and a low growth ($g_{i,t} < c \cdot ROA_{i,t}$) group, assuming c is equal to 1, and empirically test the relationship between the payout ratio and risks for high growth firms and low growth firms. Based on the moving estimates process developed by Chow (1960), Hansen (1996, 1999, and 2000), and Zeileis et al. (2002), we here try to estimate the empirical breakpoint of the structural change and examine whether the empirical breakpoint is different from our theoretical breakpoint or not.

By using the moving estimates process, we can obtain an empirical estimate of $c = 0.93$ when using beta risk and $c = 0.97$ when using total risk.¹⁸ That is, the breakpoint of the structural change is at $g_{i,t} = 0.93 \times ROA_{i,t}$ or $g_{i,t} = 0.97 \times ROA_{i,t}$. Then we redo *Model (5) to Model (8)* of Table 4 by using the empirical structural change point instead of the theoretical structural change point. In Table 5, we compare to the theoretical model and find that the estimates and significances using the empirical breakpoint of the structural change are almost same as those using the theoretical breakpoint indicated in Table 4. Results obtained from the moving estimates process show the existence of an empirical breakpoint for the relationship between the payout ratio and risks and also confirm that the dummy variable used in Eq. (36) is both theoretically and empirically acceptable.

Table 5 presents results from fixed-effects regressions of dividend payout ratios on firm characteristics for 2,645 firms during the period from 1969 to 2009. The regressions are as follows:

¹⁸ Gujarati (2009) shows that this kind of problem can be regarded as piecewise regression by using moving estimates processes. Please see the Appendix D for the details of the moving estimates process.

Table 5 Fixed-effects regressions using empirical structural change point

Model	Intercept	Beta	D*beta	Growth* beta	Total risk	D*total risk	Growth* total risk	Growth	ln (Size)	ROA	Adj- R2	F-test (p-value)
(1) $c = 0.93$ 0.931	0.29*** (2.58)	-0.21*** (-12.57)	0.77*** (58.41)					-0.02*** (-3.46)	0.005 (0.55)	-11.63*** (-58.46)	0.2891	<0.01
(2) $c = 0.97$	0.47*** (4.15)				-27.03*** (-23.53)	66.22*** (52.43)		-0.02*** (-3.50)	0.002 (0.25)	-11.50*** (-57.13)	0.2758	<0.01
(3) $c = 0.93$	0.28*** (2.56)	-0.20*** (-11.39)	0.77*** (57.82)	-0.13*** (-4.19)				0.04*** (2.60)	0.004 (0.44)	-11.51*** (-57.33)	0.2896	<0.01
(4) $c = 0.97$	0.46*** (4.10)				-28.67*** (-22.76)	66.52*** (52.53)	13.00*** (3.17)	-0.15*** (-3.60)	0.004 (0.45)	-11.56*** (-57.17)	0.2761	<0.01

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively

$$\ln \left(\frac{\text{payout ratio}_{i,t}}{1 - (\text{payout ratio}_{i,t})} \right) = \alpha + \beta_1 \text{Risk}_{i,t} + \beta_2 D(g_{i,t} < c \cdot \text{ROA}_{i,t}) \cdot \text{Risk}_i \\ + \beta_3 \text{Growth}_{i,t} + \beta_4 \text{Risk}_{i,t} \times \text{Growth}_{i,t} \\ + \beta_5 \ln(\text{Size})_{i,t} + \beta_6 \text{ROA}_{i,t} + e_{i,t}.$$

The dependent variable is the payout ratio with a logistic transformation. Based upon results from the moving estimates process, c is equal to 0.93 when using beta risk and 0.97 when using total risk. The dummy variable is equal to 1 if a firm's 5-year average growth rate is less than its 5-year average ROA, and 0 otherwise. The independent variables are beta risk (total risk), dummy times beta risk (total risk), interaction between growth rate and beta risk (total risk), growth rate, log of size, and the rate of return on assets. t -statistics are presented in parentheses, and the last column presents p -values of the F -test on the null hypothesis that the restricted *Models (1)–(4)* in Table 4 are not different from the restricted *Models (1)–(4)* in Table 5.

101.9 Interaction Among Investment, Financing, and Dividend Policies

In previous sections, we have theoretically and empirically shown that dividend policy decision will affect the value of a firm. Fama (1974) investigated the relationship between investment and dividend policy and Fama and French (2002) investigated the relationship between dividend policy and financing policy. By updating and generalizing Fama and Fama and French's models, Lee et al. (2016) investigated the relationship among investment, dividend, and financing policy.

Using the data from 1965 to 2019, we update Lee et al. (2016) paper. The main empirical results in terms of updated data are presented in Table 6. We use a three-equation system, which is presented in Eqs. 37, 38, and 39, to perform the empirical study. The empirical results of their equation system is presented in Table 6. The exogenous variables used in Table 6 are now discussed in some detail. In addition to lag-terms of the three policies, they follow Fama (1974) to respectively incorporate sales plus change in inventories (Q_{it}) and net income minus preferred dividends (P_{it}) into investment and dividend decisions. Moreover, Lee et al. (2016) follow Fama and French (2002) to add natural logarithm of lagged total assets ($\ln A_{i,t-1}$) and the lag of earnings before interest and taxes divided by total assets ($E_{i,t-1}/A_{i,t-1}$) as the determinants of leverage. Finally, Leary and Roberts (2014) argue that the characteristic of peer firms is important to influence firms' capital structure and thus they also follow them to consider lagged industry averages book leverage ($\text{Industry}_{i,t-1}$) into their leverage decision.

$$\text{Inv}_{it} = \alpha_{1i} + \alpha_{2i} \text{Div}_{it} + \alpha_{3i} \text{Leverage}_{it} + \alpha_{4i} \text{Inv}_{i,t-1} + \alpha_{5i} Q_{it} + \epsilon_{it}, \quad (37)$$

Table 6 Results of 2SLS

Dependent Variables	Independent Variables	Inv_{it}		Div_{it}		$Leverage_{it}$	
<i>Constant</i>		0.5082		0.1660***		0.1660***	
		(1.36)		(12.47)		(12.47)	
$Inv_{i,t-1}$		0.4542***	[+]				
		(52.61)					
$Div_{i,t-1}$				0.5580***	[+]		
				(52.56)			
$Leverage_{i,t-1}$						0.5645***	[+]
						(67.41)	
Inv_{it}				0.0106***	[?]	0.0138***	[+]
				(5.22)		(5.66)	
Div_{it}		21.5010	[?]			5.1026	[?]
		(0.10)				(0.16)	
$Leverage_{it}$		0.1728	[+]	-0.3060***	[?]		
		(0.29)		(-9.28)			
Q_{it}		0.1891***	[+]				
		(20.51)					
P_{it}				0.0479***	[+]		
				(18.53)			
$\ln A_{i,t-1}$						0.0043**	[+]
						(2.25)	
$E_{i,t-1}/A_{i,t-1}$						-0.1347***	[-]
						(-7.84)	
$Industry_{i,t-1}$						-0.0040	[+]
						(-0.35)	
Adjusted R-squares		0.7724		0.6493		0.4113	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$$Div_{it} = \beta_{1i} + \beta_{2i}Inv_{it} + \beta_{3i}Leverage_{it} + \beta_{4i}Div_{i,t-1} + \beta_{5i}P_{it} + \eta_{it}, \tag{38}$$

$$Leverage_{it} = \gamma_{1i} + \gamma_{2i}Inv_{it} + \gamma_{3i}Div_{it} + \gamma_{4i}Leverage_{i,t-1} + \gamma_{5i} \ln A_{i,t-1} + \gamma_{6i}(E_{i,t-1}/A_{i,t-1}) + \gamma_{7i}Industry_{i,t-1} + \xi_{it} \tag{39}$$

Table 6 shows 2SLS regression results for simultaneous-equations model. Averages of individual firms' coefficient estimates of the investment, dividend, and debt financing regressions are presented. The results of relations among these three financial decisions could be directly obtained in this method. First, the positive coefficient of $Leverage_{it}$ (which is marginally significant in investment decision), and Inv_{it} (in debt financing decision) imply that firms with higher investment have higher debt financing and vice versa. Ross (1977) and Myers and Majluf (1984)

suggest that debt is preferred to equity for managers to signal the optimistic investment opportunity to investors because firms are expected to have higher future cash flows to repay the debt. Harris and Raviv (1990) argue that debt is a device to solve the asymmetric information for the investors, because it helps to monitor managers and force the firm to liquidation. In addition, our finding that increases in debt financing enhance the funds available to outlays for investment is consistent with McDonald et al. (1975), McCabe (1979), Peterson and Benesh (1983), John and Nachman (1985), and Froot et al. (1993). Thus, our optimal debt ratio may be the result of a trade-off between the value of information (from more debt) and the cost of monitor.

Second, the significant coefficients of Div_{it} and Inv_{it} imply that dividend outlays influence investment decisions and vice versa. This finding implies that the firms may use dividend payout to signal the growth opportunity and then these firms increase their investment. The firm with higher investment input may experience the higher earnings and thus could increase the dividend payout. Such finding, that dividend payout responds to investment, confirms the model prediction of Lambrecht and Myers (2012). The relationship between dividend payout and investment is also consistent with the signaling cash flow hypothesis of dividend payout in Yoon and Starks (1995).

Third, the impact of debt financing on dividend is significantly negative, showing that firms pay more dividend when they have lower leverage level. The pecking order hypothesis of Myers and Majluf (1984) implies that firms with less leverage level have more internal funds and thus these firms tend to have capability to pay dividend. Thus, our finding seems to support the concept of Myers and Majluf (1984) and is also consistent with Jensen et al. (1992) and Fama and French (2002).

Table 6 presents the 2SLS regression results of a simultaneous equation system model for investment, dividend, and debt financing.

The coefficients are shown in averages across the 563 firms. Regressions are based on non-missing observations and winsorization at the 1st and 99th percentiles. The three endogenous variables are Inv_{it} , Div_{it} , and $Leverage_{it}$, which are net plant and equipment, dividends, and book leverage ratio, respectively. The independent variables in the investment regression are lagged investment ($Inv_{i,t-1}$), and sales plus change in inventories (Q_{it}). The independent variables in the dividend regression are lagged dividends ($Div_{i,t-1}$), and net income minus preferred dividends (P_{it}). All the variables in both of investment and dividend equations are measured on a per share basis. The independent variables in the debt financing regression are lagged book leverage ($Leverage_{i,t-1}$), natural logarithm of lagged total assets ($\ln A_{i,t-1}$), the lag of earnings before interest and taxes divided by total assets ($E_{i,t-1}/A_{i,t-1}$), and lagged industry averages book leverage ($Industry_{i,t-1}$). Numbers in the parentheses are t-statistics. The sign in bracket is the expected sign of each variable of regressions.

101.10 Summary and Concluding Remarks

In this chapter, we extend the model developed in earlier studies and improve upon the results obtained in those studies on optimal dividend payout. By allowing partial dividend payout for the consideration of flexibility, we theoretically show the

existence of the optimal dividend payout. The results obtained in this chapter are different from those of M&M (1961) because, contrary to their model, the model developed here is dynamic, is under uncertainty, and allows the firm to retain some earnings for positive NPV projects, following DeAngelo and DeAngelo (2006) and Blau and Fuller (2008). The dynamic stochastic model developed and empirically tested here is comprehensive and for the very first time includes several variables simultaneously that have not been included in any of the earlier studies. More specifically, the optimization model developed here includes, among other things, a stochastic rate of return on assets, corporate growth rate, and total risk (broken down into its systematic and firm-specific components).

We further analyze the effects of different parameters on the optimal payout ratio. A sensitivity analysis and an approximation form show a negative but nonlinear relationship between the optimal dividend payout ratio and the growth rate. We also explicitly derive the theoretical relationship between the optimal payout ratio and risks. Results of the comparative static analyses cast a different light on the relationship between corporate dividend policy and risk. That the risk and dividend payout are negatively related is generally believed. On the contrary, the relationship between risk and dividend payout is positive when the growth rate is less than the company's rate of return on assets. Only when the growth rate exceeds the rate of return on assets does the relationship between dividend payout and risk turns negative. In sum, this chapter shows for the first time that the relationship between growth and dividend payout is nonlinear. In addition, no research has ever shown that the relationship between risk rates and the dividend payout is nonlinear and influenced by other important variables, those of the company's rate of return on assets and growth rate.

Based upon our theoretical model, we develop three testable hypotheses and try to reconcile conflicts between the flexibility hypothesis and the free cash flow hypothesis in existing literature. The empirical research covers 40 years of the US data from 1969 to 2009 for 28,333 dividend-paying firm-years. Our empirical results show that the optimal dividend payout ratio is negatively, but not linearly, related to the growth rate. In addition, the optimal dividend payout ratio is negatively (positively) related to both total risk and systematic risk when the growth rate is higher (lower) than the rate of return on assets. Such results also confirm the argument that high growth firms pay dividends due to flexibility considerations and low growth firms pay dividends due to the consideration of the free cash flow problem. This result implies that dividend policy affects the value of a firm. Based upon this result, we empirically analyze interaction among investment, financing, and dividend policy.

In sum, this chapter develops a dynamic stochastic model deriving relationships between corporate dividend payout and several important financial variables. The relationships derived from the model are then tested by extensive empirical research. The relationships found between dividend payout and the growth rate and between dividend payout and risk run counter to those observed in earlier studies and also counter to generally held beliefs. Extensive empirical research validated the conclusions derived from the dynamic stochastic optimization model developed in this chapter. For the first time we report here in the literature that these relationships are

nonlinear and not solely negative as found in other studies. Thus this chapter extends the model developed in earlier studies and improves upon the results obtained in those studies on optimal dividend payout. We believe that this chapter has shown that dividend policy is relevant instead of irrelevant in terms of affecting the value of a firm.

Appendix A. Derivation of Eq. (19)

This appendix presents a detailed derivation of the solution to the variable partial differential equation, Eq. (17), which is similar to Gould's (1968) Eq. (9) in investigation the adjustment cost. Following Gould's (1968) approach, we first derive a general solution for a standard variable partial differential equation. Then we apply this general equation to solve Eq. (17). The standard variable partial differential equation can be defined as

$$\dot{p}(t) + g(t)p(t) = q(t). \quad (40)$$

As a particular case of Eq. (40), the equation

$$\dot{p}(t) + g(t)p(t) = 0 \text{ or } \frac{\dot{p}(t)}{p(t)} = -g(t) \quad (41)$$

has a solution

$$p(t) = c \cdot \exp\left(-\int g(t)dt\right). \quad (42)$$

By substituting constant c with function $c(t)$, we have the potential solution to Eq. (40)

$$p(t) = c(t) \cdot \exp\left(-\int g(t)dt\right). \quad (43)$$

Taking a differential with respect to t , we obtain

$$\begin{aligned} \dot{p}(t) &= \dot{c}(t) \cdot \exp\left(-\int g(t)dt\right) - c(t) \cdot \exp\left(-\int g(t)dt\right)g(t) \\ &= \dot{c}(t) \cdot \exp\left(-\int g(t)dt\right) - p(t)g(t). \end{aligned} \quad (44)$$

Therefore,

$$\dot{p}(t) + p(t)g(t) = \dot{c}(t) \cdot \exp\left(-\int g(t)dt\right). \quad (45)$$

From Eq. (40) and Eq. (45), we have

$$\dot{c}(t) \cdot \exp \left(- \int g(t) dt \right) = q(t). \quad (46)$$

Equivalently,

$$\dot{c}(t) = q(t) \cdot \exp \left(\int g(t) dt \right). \quad (47)$$

Therefore,

$$c(t) = \int q(t) \cdot \exp \left(\int g(t) dt \right) dt. \quad (48)$$

Substituting Eq. (48) into Eq. (42), we have the general solution of Eq. (40),

$$p(t) = \exp \left(- \int g(t) dt \right) \cdot \left[\int q(t) \exp \left(\int g(t) dt \right) dt \right]. \quad (49)$$

To solve Eq. (17), we will apply the above result. Let $g(t) = \delta \frac{\dot{m}(t)}{m(t)} - k$ and $q(t) = -G(t)$.

Because

$$\begin{aligned} \exp \left(\int g(t) dt \right) &= \exp \left(\int \left(\delta \frac{\dot{m}(t)}{m(t)} - k \right) dt \right) \\ &= \exp \left(\delta \int \frac{\dot{m}(t)}{m(t)} dt - kt \right) \\ &= \exp \left(\delta \ln(m(t)) - kt + c \right) \\ &= c_1 \cdot m(t)^\delta \exp(-kt), \end{aligned} \quad (50)$$

where $c_1 > 0$,

then we have

$$P(t) = c_2 \cdot m(t)^{-\delta} \exp(kt) \cdot \left[\int q(t) c_3 \cdot m(t)^\delta \exp(-kt) dt \right], \quad (51)$$

where $c_2 > 0$, and $c_3 > 0$

or equivalently,

$$P(t) = c_4 \cdot \frac{e^{kt}}{m(t)^\delta} \cdot \left[\int -G(t) \frac{m(t)^\delta}{e^{kt}} dt \right], \quad (52)$$

where $c_4 > 0$.

Finally, we have

$$P(t) = c \frac{e^{kt}}{m(t)^k} \cdot \int G(t)m(t)^\delta e^{-kt} dt. \quad (53)$$

Changing from an indefinite integral to a definite integral, Eq. (52) can be shown as

$$p(t) = \frac{e^{kt}}{m(t)^\delta} \int_t^T G(s)m(s)^\delta e^{-ks} ds,$$

which is Eq. (19).

Appendix B. Derivation of Eq. (21)

This appendix presents a detailed derivation of Eq. (21). In Eq. (20), the initial value of the firm can be expressed as

$$p(0) = \frac{1}{m(0)^\delta} \int_0^T \left\{ (a + b\bar{l} - h)A(0)e^{th}m(t)^{\delta-1} - \alpha'A(0)^2\rho(t)^2\sigma(t)^2e^{2th}m(t)^{\delta-2} \right\} e^{-kt} dt. \quad (20)$$

To maximize firm value, the number of shares outstanding at each point of time should be determined; therefore, the objective function can be written as follows:

$$\max_{\{m(t)\}_{t=0}^T} p(0). \quad (54)$$

Following the Euler-Lagrange condition (see Chiang 1984), we take first order conditions on the objective function with respect to $m(t)$, where $t \in [0, T]$ and let such first order conditions be equal to zero.

$$\frac{1}{m(0)^\delta} \left\{ (\delta - 1)(a + b\bar{l} - h)A(0)e^{th}m(t)^{\delta-2} - \alpha'A(0)^2\rho(t)^2\sigma(t)^2e^{2th}m(t)^{\delta-3}(\delta - 2) \right\} e^{-kt} dt = 0, \quad (55)$$

where $t \in [0, T]$.

To simplify Eq. (55),

$$\begin{aligned} & (\delta - 1)(a + b\bar{l} - h)A(0)e^{th}m(t)^{\delta-2} - \alpha'A(0)^2\rho(t)^2\sigma(t)^2e^{2th}m(t)^{\delta-3}(\delta - 2) \\ & = 0, \end{aligned} \quad (56)$$

where $t \in [0, T]$, which is Eq. (21).

Appendix C. Derivation of Eqs. (28) and (29)

This appendix presents a detailed derivation of both Eq. (28) and Eq. (29). In Eq. (35), the optimal payout ratio with no changes in total risk or systematic risk is

$$[\overline{D}(t)/\overline{x}(t)] = \left(1 - \frac{h}{a + b\overline{l}}\right) \left[\frac{-k + he^{(h-k)(T-t)}}{h-k}\right]. \quad (35)$$

Considering the finite growth case, if $(h-k)(T-t) < 1$, then following MacLaurin expansion, the $e^{(h-k)(T-t)}$ can be expressed as

$$\begin{aligned} e^{(h-k)(T-t)} &= 1 + (h-k)(T-t) + \frac{(h-k)^2(T-t)^2}{2!} + \frac{(h-k)^3(T-t)^3}{3!} \\ &\approx 1 + (h-k)(T-t) \end{aligned} \quad (57)$$

Therefore, Eq. (31) can be approximately written as

$$[\overline{D}(t)/\overline{x}(t)] \approx \left(1 - \frac{h}{a + b\overline{l}}\right) (1 + h(T-t)),$$

which is Eq. (38).

We further take the partial derivative of Eq. (33) with respect to the growth rate. Then the partial derivative of optimal payout ratio with respect to the growth rate can be approximately written as

$$\frac{\partial [\overline{D}(t)/\overline{x}(t)]}{\partial h} \approx \left(\frac{(a + b\overline{l})(T-t) - 2h(T-t) - 1}{a + b\overline{l}}\right),$$

which is Eq. (29).

Appendix D. Using Moving Estimates Process to Find the Structural Change Point in Eq. (36)

To estimate the empirical breakpoint, we first assume the dummy variable associated with risk as $D(g_{i,t} < c \cdot ROA_{i,t})$. We introduce a no structural change model, Eq. (58), and a structural change model, Eq. (59). In Eq. (59), c is a continuous constant variable, ranging from 0.2 to 1.8 ($c \in [0.2, 1.8]$). The breakpoint of structural change is at $g_{i,t} = c \cdot ROA_{i,t}$.

$$\begin{aligned} \ln \left(\frac{\text{payout ratio}_{i,t}}{1 - (\text{payout ratio}_{i,t})} \right) &= \alpha + \beta_1 \text{Risk}_{i,t} + \beta_2 \text{Growth}_{i,t} + \beta_3 \ln(\text{Size})_{i,t} \\ &\quad + \beta_4 ROA_{i,t} + e_{i,t} \end{aligned} \quad (58)$$

$$\ln \left(\frac{\text{payout ratio}_{i,t}}{1 - (\text{payout ratio}_{i,t})} \right) = \alpha + \beta'_1 \text{Risk}_{i,t} + \beta'_2 D(g_{i,t} < c \cdot ROA_{i,t}) \cdot \text{Risk}_i + \beta'_3 \text{Growth}_{i,t} + \beta'_4 \ln(\text{Size})_{i,t} + \beta'_5 ROA_{i,t} + e_{i,t} \tag{59}$$

By using the moving estimates process, we calculate the F statistics for all potential structural change points between $g_{i,t} = 0.2 \times ROA_{i,t}$ and $g_{i,t} = 1.8 \times ROA_{i,t}$. The F test is under the null hypothesis that no structural change on the relationship between the payout ratio and the risk. That is, Eq. (58) is identical to Eq. (59). Finally, we can locate the breakpoint of the structural change at the point with highest value of F -statistics.

From Table 7, the process has a clear peak at $c = 0.93$ when using the beta risk as the independent variable and $c = 0.96$ when using the total risk as the independent variable. Figure 2 also graphically presents a peak F -statistics point at $c = 0.93$

Table 7 Moving estimates process for testing the break point of structural change

F-statistics			F-statistics		
c	Beta risk	Total risk	c	Beta risk	Total risk
0.20	120.38	92.78	1.01	658.58	533.26
0.25	147.77	114.94	1.02	646.43	524.02
0.30	205.83	153.89	1.03	636.99	516.94
0.35	252.56	179.52	1.04	621.57	506.64
0.40	333.30	244.85	1.05	619.94	507.41
0.45	395.33	298.05	1.06	610.91	503.46
0.50	421.89	320.42	1.07	616.34	505.71
0.55	445.75	340.36	1.08	610.81	502.08
0.60	505.14	213.02	1.09	608.96	505.16
0.65	532.57	205.06	1.10	609.88	496.72
0.70	572.82	411.85	1.15	610.61	499.96
0.75	600.57	452.04	1.20	575.99	473.31
0.80	633.15	485.64	1.25	543.25	445.85
0.85	674.48	524.66	1.30	532.27	431.02
0.90	675.53	542.20	1.35	505.57	408.10
0.91	682.40	547.18	1.40	487.91	388.90
0.92	678.08	542.58	1.45	446.97	364.58
0.93	682.43	549.74	1.50	408.25	326.46
0.94	679.55	547.81	1.55	383.32	304.93
0.95	679.52	546.92	1.60	345.32	266.34
0.96	675.88	549.83	1.65	327.34	249.22
0.97	668.42	543.73	1.70	309.13	231.42
0.98	669.35	547.23	1.75	296.67	223.43
0.99	664.33	548.16	1.80	278.69	212.60
1.00	622.77	545.64			

(0.96) in terms of beta risk (total risk). Results from the moving estimates process indicate that a structural change on the relationship exists between the payout ratio and risks. The breakpoint of the structural change is at $g_{i,t} = 0.93 \times ROA_{i,t}$ or $g_{i,t} = 0.97 \times ROA_{i,t}$.

Table 7 shows the F -statistics of moving estimates processes. The nonstructural change regression and structural change regression are as follows:

$$\ln \left(\frac{\text{payout ratio}_{i,t}}{1 - (\text{payout ratio}_{i,t})} \right) = \alpha + \beta_1 \text{Risk}_{i,t} + \beta_2 \text{Growth}_{i,t} + \beta_3 \ln (\text{Size})_{i,t} + \beta_4 ROA_{i,t} + e_{i,t}$$

$$\ln \left(\frac{\text{payout ratio}_{i,t}}{1 - (\text{payout ratio}_{i,t})} \right) = \alpha + \beta'_1 \text{Risk}_{i,t} + \beta'_2 D(g_{i,t} < c \cdot ROA_{i,t}) \cdot \text{Risk}_{i,t} + \beta'_3 \text{Growth}_{i,t} + \beta'_5 \ln (\text{Size})_{i,t} + \beta'_6 ROA_{i,t} + e_{i,t}.$$

The dependent variable is the payout ratio with a logistic transformation. The breakpoint, c , is between 0.2 times and 1.8 times the rate of return on total assets. The dummy variable is equal to 1 if a firm's 5-year average growth rate is less than c times its 5-year average ROA, and 0 otherwise. The independent variables are beta risk (total risk), dummy times beta risk (total risk), growth rate, log of size, and the rate of return on assets. F -statistics are under the null hypothesis that the relationship between the payout ratio and the risk does not depend on the growth rate of a firm.

Figure 2 shows the F -statistics of moving estimates processes. The nonstructural change regression and structural change regression are as follows.

$$\ln \left(\frac{\text{payout ratio}_{i,t}}{1 - (\text{payout ratio}_{i,t})} \right) = \alpha + \beta_1 \text{Risk}_{i,t} + \beta_2 \text{Growth}_{i,t} + \beta_3 \ln (\text{Size})_{i,t} + \beta_4 ROA_{i,t} + e_{i,t}$$

$$\ln \left(\frac{\text{payout ratio}_{i,t}}{1 - (\text{payout ratio}_{i,t})} \right) = \alpha + \beta'_1 \text{Risk}_{i,t} + \beta'_2 D(g_{i,t} < c \cdot ROA_{i,t}) \cdot \text{Risk}_{i,t} + \beta'_3 \text{Growth}_{i,t} + \beta'_5 \ln (\text{Size})_{i,t} + \beta'_6 ROA_{i,t} + e_{i,t}$$

The dependent variable is the payout ratio with a logistic transformation. The breakpoint, c , is between 0.2 times and 1.8 times rate of return on total assets. The dummy variable is equal to 1 if a firm's 5-year average growth rate is less than c times its 5-year average ROA, and 0 otherwise. The independent variables are beta risk (total risk), dummy times beta risk (total risk), growth rate, log of size, and the rate of return on assets. F -statistics is under the null hypothesis that the relationship between the payout ratio and the risk does not depend on the growth rate of a firm. The risk used in Fig. 1 is the beta, and the risk used in Fig. 2 is the total risk.

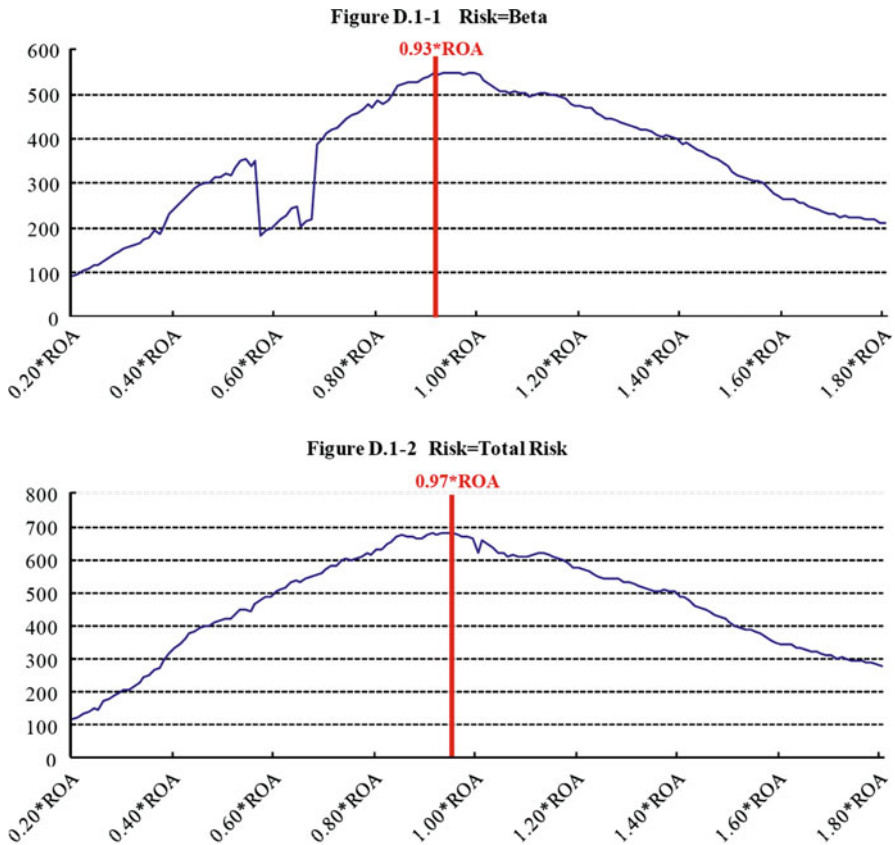


Fig. 2 Moving estimates process for testing the break point of structural change

References

- Agrawal, A., and N. Jayaraman. 1994. The dividend policies of all-equity firms: A direct test of the free cash flow theory. *Managerial and Decision Economics* 15: 139–148.
- Aivazian, V., L. Booth, and S. Cleary. 2003. Do emerging market firms follow different dividend policies from U.S. firms? *Journal of Financial Research* 26: 371–387.
- Aoki, M. 1967. *Optimization of stochastic systems*. New York: Academic.
- Asquith, P., and D.W. Mullins Jr. 1983. The impact of initiating dividend payments on shareholders' wealth. *Journal of Business* 56: 77–96.
- Bellman, R. 1990. *Adaptive control process: A guided tour*. Princeton: Princeton University Press.
- . 2003. *Dynamic programming*. Mineola: Dover Publication.
- Benartzi, S., R. Michaely, and R. Thaler. 1997. Do changes in dividends signal the future or the past? *Journal of Finance* 52: 1007–1034.
- Bernheim, B.D., and A. Wantz. 1995. A tax-based test of the dividend signaling hypothesis. *American Economic Review* 85: 532–551.

- Bhattacharya, S. 1979. Imperfect information, dividend policy, and “the bird in the hand” fallacy. *Bell Journal of Economics* 10: 259–270.
- Blau, B.M., and K.P. Fuller. 2008. Flexibility and dividends. *Journal of Corporate Finance* 14: 133–152.
- Boehmer E., C. M. Jones, and X. Zhang. 2010. Shackling short sellers: The 2008 shorting ban. Working paper.
- Brennan, M.J. 1973. An approach to the valuation of uncertain income streams. *Journal of Finance* 28: 661–674.
- . 1974. An inter-temporal approach to the optimization of dividend policy with pre-determined investments: Comment. *Journal of Finance* 29: 258–259.
- Brook, Y., W.T. Charlton Jr., and R.J. Hendershott. 1998. Do firms use dividends to signal large future cash flow increases? *Financial Management* 27: 46–57.
- Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2006. *Robust inference with multiway clustering*, Working paper.
- Chiang, A.C. 1984. *Fundamental methods of mathematical economics*. 3rd ed. New York: McGraw-Hill.
- Chow, G.C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28: 591–605.
- DeAngelo, H., and L. DeAngelo. 2006. The irrelevance of the MM dividend irrelevance theorem. *Journal of Financial Economics* 79: 293–315.
- DeAngelo, H., L. DeAngelo, and D.J. Skinner. 1996. Reversal of fortune: Dividend signaling and the disappearance of sustained earnings growth. *Journal of Financial Economics* 40: 341–371.
- DeAngelo, H., L. DeAngelo, and R.M. Stulz. 2006. Dividend policy and the earned/contributed capital mix: A test of the life-cycle theory. *Journal of Financial Economics* 81: 227–254.
- Denis, D.J., and I. Osobov. 2008. Why do firms pay dividends? International evidence on the determinants of dividend policy. *Journal of Financial Economics* 89: 62–82.
- Denis, D.J., D.K. Denis, and A. Sarin. 1994. The information content of dividend changes: Cash flow signaling, overinvestment, and dividend clienteles. *Journal of Financial and Quantitative Analysis* 29: 567–587.
- Easterbrook, F.H. 1984. Two agency-cost explanations of dividends. *American Economic Review* 74: 650–659.
- Fama, E.F. 1974. The empirical relationships between the dividend and investment decisions of firms. *The American Economic Review* 64: 304–318.
- Fama, E.F., and K.R. French. 2001. Disappearing dividends: Changing firm characteristics or lower propensity to pay? *Journal of Financial Economics* 60: 3–43.
- . 2002. Testing trade-off and pecking order predictions about dividends and debt. *The Review of Finance Studies* 15: 1–33.
- Fenn, G.W., and N. Liang. 2001. Corporate payout policy and managerial stock incentives. *Journal of Financial Economics* 60: 45–72.
- Froot, K.A., D.S. Scharfstein, and J.C. Stein. 1993. Risk management: Coordinating corporate investment and financing policies. *The Journal of Finance* 48: 1629–1658.
- Gabudean, R. C. 2007. *Strategic interaction and the co-determination of firms financial policies*, Working Paper.
- Gordon, M.J. 1962. The savings investment and valuation of a corporation. *Review of Economics and Statistics* 44: 37–51.
- Gould, J.P. 1968. Adjustment costs in the theory of investment of the firm. *Review of Economic Studies* 35: 47–55.
- Grullon, G., R. Michaely, and B. Swaminathan. 2002. Are dividend changes a sign of firm maturity? *Journal of Business* 75: 387–424.
- Gujarati, D.N. 2009. *Basic econometrics*. 5th ed. New York: McGraw-Hill.
- Hamilton, J.D. 1994. *Time series analysis*. 1st ed. Princeton: Princeton University Press.
- Hansen, B.E. 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64: 413–430.

- . 1999. Threshold effects in nondynamic panels: Estimation, testing, and inference. *Journal of Econometrics* 93: 345–368.
- . 2000. Sample splitting and threshold estimation. *Econometrica* 68: 575–603.
- Harris, M., and A. Raviv. 1990. Capital structure and the informational role of debt. *The Journal of Finance* 45: 321–349.
- Healy, P.M., and K.G. Palepu. 1988. Earnings information conveyed by dividend initiations and omissions. *Journal of Financial Economics* 21: 149–175.
- Higgins, R.C. 1977. How much growth can a firm afford? *Financial Management* 6: 7–16.
- Hogg, R.V., and A.T. Craig. 2004. *Introduction to mathematical statistics*. 6th ed. Englewood Cliffs: Prentice Hall.
- Howe, K.M., J. He, and G.W. Kao. 1992. One-time cash flow cash and free cash-flow theory: Share repurchases and special dividends. *Journal of Finance* 47: 1963–1975.
- Intriligator, M.D. 2002. *Mathematical optimization and economic theory*. Philadelphia: Society for Industrial and Applied Mathematics.
- Jagannathan, M., C.P. Stephens, and M.S. Weisbach. 2000. Financial flexibility and the choice between dividends and stock repurchases. *Journal of Financial Economics* 57: 355–384.
- Jensen, M.C. 1986. Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review* 76: 323–329.
- Jensen, G.R., D.P. Solberg, and T.S. Zorn. 1992. Simultaneous determination of insider ownership, debt, and dividend policies. *Journal of Financial and Quantitative Analysis* 27: 247–263.
- John, K., and D.C. Nachman. 1985. Risky debt, investment incentives, and reputation in a sequential equilibrium. *The Journal of Finance* 40: 863–878.
- John, K., and J. Williams. 1985. Dividends, dilution, and taxes: A signaling equilibrium. *Journal of Finance* 40: 1053–1070.
- Kalay, A., and U. Loewenstein. 1986. The informational content of the timing of dividend announcements. *Journal of Financial Economics* 16: 373–388.
- Kalay, A., and R. Michaely. 1993. *Dividends and taxes: A reexamination*, Working Paper.
- Kao, C., and C. Wu. 1994. Tests of dividend signaling using the March-Merton model: A generalized friction approach. *Journal of Business* 67: 45–68.
- Kreyszig, E. 2010. *Advanced engineering mathematics*. 10th ed. New York: Wiley.
- Lambrecht, B.M., and S.C. Myers. 2012. A Lintner model of payout and managerial rents. *Journal of Finance* 67: 1761–1810.
- Lang, L.H.P., and R.H. Litzenberger. 1989. Dividend announcements: Cash flow signaling vs. free cash flow hypothesis? *Journal of Financial Economics* 24: 181–191.
- Leary, M.T., and M.R. Roberts. 2014. Do peer firms affect corporate financial policy? *The Journal of Finance* 69: 139–178.
- Lee, Cheng F., and A. Lee. 2021. Investment, financing, dividend, and production policies: Review and integration. Review of Pacific Basin Financial Market and Policy, Forthcoming.
- Lee, Cheng F., and J. Shi. 2010. Application of alternative ODE in finance and econometric research. In *Handbook of quantitative finance and risk management*, ed. Cheng F. Lee, Alice C. Lee, and John Lee, 1293–1300. Springer. New York.
- Lee, Cheng-Few, Woan-Lih Liang, Fu-Lai Lin, and Yating Yang. 2016. Application of simultaneous equation in finance research: Methods and empirical results. *Review of Quantitative Finance and Accounting* 47 (4): 943–971.
- Lee, Cheng F., and A. Lee. 2021. Investment, financing, dividend, and production policies: Review and integration. Review of Pacific Basin Financial Market and Policy 24: 2150018.
- Lerner, E.M., and W.T. Carleton. 1966a. Financing decisions of the firm. *Journal of Finance* 21: 202–214.
- . 1966b. *A theory of financial analysis*. New York: Harcourt, Brace & World.
- Lie, E., 2005. Operating performance following dividend decreases and omissions. *Journal of Corporate Finance* 12, 27–53.
- Lintner, J. 1962. Dividends, earnings, leverage, stock prices and the supply of capital to corporations. *Review of Economics and Statistics* 44: 243–269.

- . 1963. The cost of capital and optimal financing of corporate growth. *Journal of Finance* 18: 292–310.
- . 1964. Optimal dividends and corporate growth under uncertainty. *Quarterly Journal of Economics* 78: 49–95.
- . 1965. Security prices, risk, and maximal gains from diversification. *Journal of Finance* 20: 587–615.
- McCabe, G.M. 1979. The empirical relationship between investment and financing: A new look. *Journal of Finance and Quantitative Analysis* 14: 119–135.
- McDonald, J.G., B. Jacquillat, and M. Nussenbaum. 1975. Dividend, investment and financing decisions: Empirical evidence on French firms. *Journal of Finance and Quantitative Analysis* 10: 741–755.
- Miller, M.H., and F. Modigliani. 1961. Dividend policy, growth, and the valuation of shares. *Journal of Business* 34: 411–433.
- Miller, M.H., and K. Rock. 1985. Dividend policy under asymmetric information. *Journal of Finance* 40: 1031–1051.
- Mossin, J. 1966. Equilibrium in a capital asset market. *Econometrica* 34: 768–783.
- Myers, S.C., and N.S. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Finance Economics* 13: 187–221.
- Nissim, D., and A. Ziv. 2001. Dividend changes and future profitability. *Journal of Finance* 56: 2111–2133.
- Petersen, M.A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22: 435–480.
- Peterson, P.P., and G.A. Benesh. 1983. A reexamination of the empirical relationship between investment and financing decisions. *Journal of Finance and Quantitative Analysis* 18: 439–453.
- Pratt, J.W. 1964. Risk aversion in the small and in the large. *Econometrica* 32: 122–136.
- Ross, S.A. 1977. The determination of financial structure: The incentive-signaling approach. *Bell Journal of Economics* 8: 23–40.
- Rozeff, M.S. 1982. Growth, beta and agency costs as determinants of dividend payout ratios. *Journal of Financial Research* 5: 249–259.
- Sasson, B.Y., and R. Kolodny. 1976. Dividend policy and capital market theory. *Review of Economics and Statistics* 58, 181–190.
- Sharpe, W.F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
- Thompson, S. B. 2011. *Simple formulas for standard errors that cluster by both firm and time*, *Journal of Financial Economics* 99: 1–10.
- Wallingford, B.A., II. 1972a. An inter-temporal approach to the optimization of dividend policy with predetermined investments. *Journal of Finance* 27: 627–635.
- . 1972b. A correction to “an inter-temporal approach to the optimization of dividend policy with predetermined investments”. *Journal of Finance* 27: 944–945.
- Yoon, P.S., and L.T. Starks. 1995. Signaling, investment opportunities, and dividend announcements. *Review of Financial Studies* 8: 995–1018.
- Zeileis, A., F. Leisch, K. Hornik, and C. Kleiber. 2002. Strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software* 7: 1–38.



Mergers and Acquisitions: Principles and Practices

102

Harvey A. Poniachek

Abstract

Mergers and acquisitions (M&As) have been part of the corporate experience for over a century and most successful companies rely heavily on them to attain their strategic goals faster, cheaper, and less risky than through organic expansion (Poniachek, Harvey A., ed. 2019. *Mergers and acquisitions: A practitioner's guide to successful deals*. World Scientific; Bruner, Robert F. 2004. *Applied mergers and acquisitions*. University edition. Wiley Finance). M&As cover a wide range of transaction types that are complex and risky, are subject to intensive regulations related to securities issuing, antitrust, and corporate governance that often require the services of various professionals, which are predominantly procured from outside the corporation. Post-merger integration is vital, and its efficient application could assure successful deals.

The M&A market has globalized, and cross-border business have become quite significant. Covid-19 medical crisis promoted alternative deals that could achieve the same objectives with limited upfront investment and risk. M&A transactions are expected to yield synergies from reduction in cost, and expansion of sales and earnings. Expected synergies could affect the valuation, which is an art and science that applies primarily through the income and market approaches.

This chapter surveys the main theories and forms of M&As, and addresses the principles and analytical methods for decision-making. We discuss the legal framework, including issues relating to antitrust, corporate governance, and anti-takeover defenses, accounting, leverage buyouts (LBOs), payment methods and their implications, valuation, and cross-border deals.

H. A. Poniachek (✉)

Department of Finance and Economics, Rutgers University, Newark, NJ, USA

e-mail: harvey.poniachek@busiess.rutgers.edu

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_105

2457

Keywords

Mergers and acquisitions · Strategic goals · M&A transactions · Antitrust · Corporate governance · Post-merger integration · Synergies · Cross-border business · Covid-19 · Valuation · Fairness opinion · Leverage buy out · Anti-takeover defenses

JEL Classification

G1 · G3 · L1 · L2 · L4

102.1 Introduction

Mergers and acquisitions (M&As) have been part of the corporate experience for over a century and virtually every company has undertaken major deals with significantly business implications. Build versus buy analysis assists firms to determine whether to pursue organic development or M&A deals. The world's most successful companies rely heavily on mergers and acquisitions to attain their strategic goals faster, cheaper, and less risky than through organic expansion (Bruner 2004). Investments via mergers and acquisitions have become essential for corporations seeking to enhance competitiveness, achieve diversification, access emerging technologies, take advantage of favorable changes in economic and the regulatory environment, and promote supply chain participation in globalization. Expansionary investment decisions via M&As involve acquisition of other firms through traditional methods, including mergers, consolidation, and purchase of assets (Poniachek 2019).

M&As are important corporate events that allow companies to achieve a wide variety of strategic goals, including expansion into new markets (horizontal integration), build a more efficient supply chain (vertical integration), mollify competition, achieve operational efficiencies, or diversify business-portfolio risk (Schill and Loutskina 2017). External investment through M&A and internal development are complementary and facilitate entry into new markets and businesses (Weston et al. 2004). Post-merger integration (PMI) could affect the merger process success or failure.

The M&A market has globalized, and cross-border business have become quite significant. Covid-19 crisis disrupted global supply chains (EY 2020) and reduced M&As deals, but an expected upturn in economic growth coupled with significant cash balances held by the US corporate sector and private equity (PE) firms could provide a strong impetus for deal activities going forward. Over 60% of corporate executives surveyed by Deloitte in August 2020 expect M&A activity to return to pre-Covid-19 levels within a year. However, past M&A experience (PWC 2020) might not be fully applicable in the future because deals will be smaller, more deals will include earnout provisions that could help reconcile valuation differences, and dealmakers will need to incorporate a higher risk premium into the discount rate.

The challenging business circumstances due to Covid-19 medical crisis promoted alternative deals, including joint ventures, alliances, and partnerships as alternatives to traditional investment approaches (Elie et al. 2016). Alternative deals could be used to achieve the same strategic objectives as traditional M&As by offering access to resources and markets that are otherwise unavailable with limited upfront investment, risk mitigation through partnership, and accelerated time to market. In alternative deals companies acquire minority stakes or establish cooperative arrangements (Kengelbach et al. 2020) as an effective approach to pursue strategic goals at reduced risk. Alternative deals and classic M&A represent significantly different deal structures, but have similar success rates, with approximately 40% of alternative deals failing to achieve their stated financial and/or strategic goals, and 60% were successful (Kengelbach et al. 2020). In 2019, there were 11,000 joint ventures and business alliances driven largely by software and IT services, commercial and professional services in the Asia-Pacific and in North America.

Special Purpose Acquisition Companies (SPACs), also referred to as “blank check,” are shell corporations that raise funds from sponsors and investors, go public through a simplified IPO process, and then merge with target firms (Surance et al. 2021; Layne et al. 2018). SPACs like entities existed for decades in various forms (Malmberg et al. 2020), but their numbers have increased steadily from 2013 and reached a record in 2020, and 200 SPACs have raised \$80 billion and accounted for 50% of the IPO during that year (Harroch et al. 2020). Their growth is attributed to competitive IPO pricing and a record influx of high-profile investors and management teams, which turned SPACs into significant players for potential sponsors, investors, and target operating companies. Foreign businesses deem publicly listed SPACs attractive merger partners.

A SPAC’s life begins with its initial formation, followed by an IPO, search for an acquisition target, and finally closing of an acquisition. The funds raised in the IPO are placed in a trust account while the SPAC seeks to acquire an existing operating company within 18–24 months, but if no acquisition occurs in the above timeframe, the cash is returned to its investors. The SPAC sponsor usually purchases founder shares prior to the SPAC IPO filing that comprise of one share of common stock, plus a warrant to acquire common stock at a price of \$10 and following a merger their shares are converted into public shares. If the SPAC successfully completes an acquisition, the target becomes a public company, otherwise the cash held in its trust account is returned to its investors, unless if the SPAC extends its timeline via a proxy process.

M&A transactions require a multitude of participants with different specialization and responsibilities, including the corporate executives and board, investment bankers, accountants, tax specialists, and attorneys. The merger process commonly commences with the formulation of the investment objectives, followed by identification of the target, negotiations, due diligence, valuation, reaching an agreement, address funding and taxation implications, determine the transaction structure, and protect against various risks. The transactors need to obtain antitrust approval, assure shareholders’ approval, and apply effective post-merger integration.

M&As need to meet intensive regulations by the US Federal Government (SEC, FTC/DOJ) and States' corporation laws, and transactions are subject to hostile deals against which companies have developed effective defenses. Corporate governance, specifically the role of the board and the duties of directors are well established by case laws and States' corporation laws.

Divestitures remain an important feature of M&A and driven both by organizations seeking to cash in on high valuations and those aiming to reposition assets in advance of a downturn (Deloitte 2020). Valuation is an art and science and applied primarily through the income (DCF) and market (ratios/multiples) approaches that often yield results that requires investment bankers' fairness opinions to support the outcomes.

M&A transactions are complex and risky, and are affected by the state of the business cycle, financial conditions, and are legally intensive and subject to various laws and regulations, various taxation laws and accounting requirements. External investment through M&As are expected to yield synergies from reduction in cost, and lower taxes and cost of capital, and expansion of sales and earnings. In evaluating a proposed M&A, the question is whether merging parties be worth more together than separately and would their shareholders be better off in the post-merger than before? We note that a high share of M&A deals fails, but more effective management could lead to greater success.

This chapter surveys the main theories and forms of M&As, and addresses the principles and analytical methods for decision-making. We discuss the legal framework, including issues relating to antitrust, corporate governance, and anti-takeover defenses, accounting, leverage buyouts (LBOs), and payment methods and their implications.

102.2 Process and Theories, Data and Trends

Mergers and acquisitions comprise a wide range of transaction types as shown in Figure 1. Acquisitions include merger or consolidation, acquisition of stock, or acquisition of assets. Merger refers to the absorption of one firm by another where the acquired firm ceases to exist as a separate entity. A consolidation is a merger of two companies where a new entity is created and both firms cease to exist as legal entities. In an acquisition of assets, one firm acquires all or part of the assets of another firm.

A variety of mergers waves occurred since the late 1800s, shown in Table 1, that were driven primarily by favorable economic and financial conditions, deregulations of industries and favorable regulatory changes, technological innovations, favorable infrastructural changes, globalization, and the search for economies of scale and scope (Table 2).

Acquisitions are classified into three types, including

- Horizontal mergers involving two firms operating in the same or similar business activity that could provide economies of scale and scope.
- Vertical mergers between firms in different businesses, or different value-added functions, designed to internalize transactions and benefits by lowering

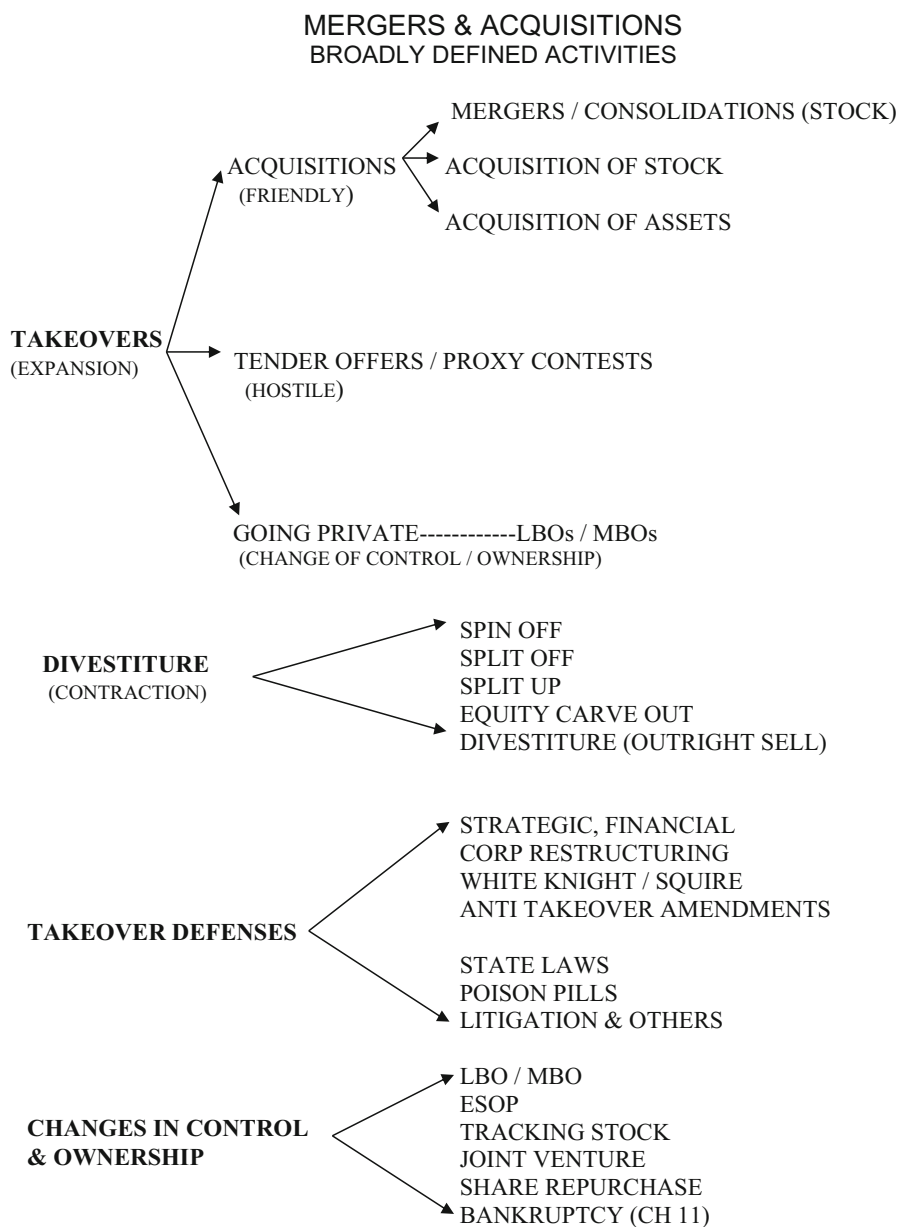


Fig. 1 Mergers and acquisitions

transaction cost through more efficient contracting, reduced or eliminated bargaining cost, harmonization of corporate objectives, reduced risk thanks to greater information availability, and adaptations to cope with regulations.

Table 1 Seven US merger waves

1. 1897–1904	Horizontal mergers
2. 1916–1929	Vertical mergers
3. 1960s	Conglomerate mergers
4. 1981–1989	Big deals decade and LBOs, hostile takeovers
5. 1992–2007	Strategic mergers
6. 2009–2019	Emergence of the hedge funds and private equity
7. 2020–	Alternative deals include alliances and joint ventures

Source: Gaughan (2011), Weston et al. (2004, p. 7 and Ch 7), and Bruner (2004)

Note: The sixth and seventh waves were defined by Harvey Poniachek

Table 2 Merger waves’ main influences

Favorable economic and financial conditions
Deregulations of major industries and favorable regulatory changes
Technological innovation
Conducive financial environment
Favorable infrastructure changes
Globalization
Industry organization
Economies of scale and scope

Source: Harvey Poniachek

- Conglomerate mergers involve firms in unrelated types of business activity that could include product extension or geographic market extension.

In 2019, mergers deals reached 49,327 valued at \$3.27 trillion as shown in Fig. 2a and b. North America (predominantly the USA) accounted for 32% of the worldwide deals and 60% of total value. The year 2020 reflected the global Covid-19 health crisis and recession, with the number of deals and their value declined to 43, 252 and \$2.6 trillion (annualized data).

102.2.1 M&A Theories/Models

Companies could invest internally or externally through mergers and acquisitions. The advantages and disadvantages of mergers versus internal development are addressed in Table 3 that favors external investment through M&As. The primary goals of M&A are to increase growth, enter a new market, gain new products and technologies, and maximize shareholder value. Several synergies, shown in Table 4, could arise from M&As that could be generated from operating and/or financial sources, and provide additional value to the combined firms (Bruner 2004; Damodaran 2001). Synergy is a major attraction of mergers, and it is defined as $2 + 2 = 5$. If an acquirer is valued at V_a and the target value is V_b , then the value of the combined firm following the merger is V_{ab} (Brealey et al. 2020):

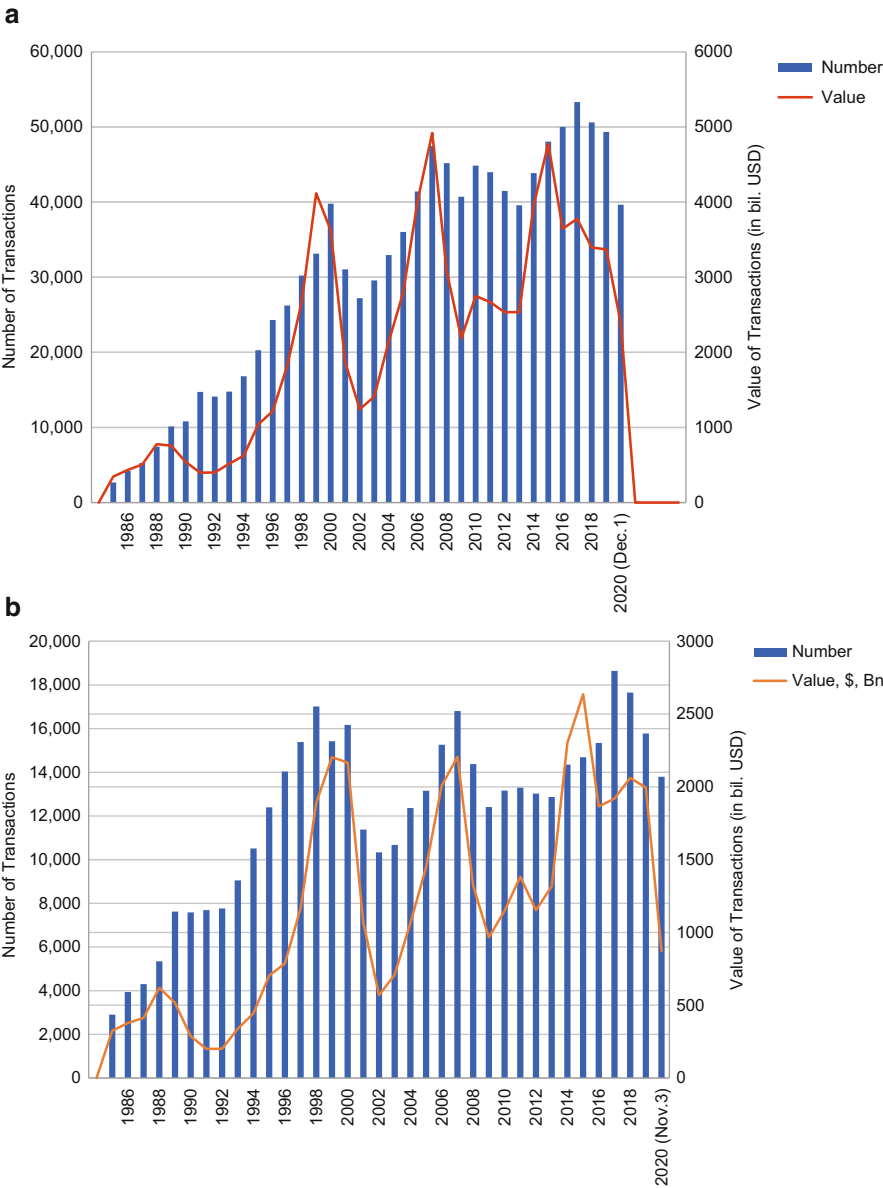


Fig. 2 (a) Mergers and acquisitions worldwide. (b) Mergers and acquisitions North America. <https://imaa-institute.org/mergers-and-acquisitions-statistics/>

$$\text{Synergy or gain } \Delta V_{ab} = V_{ab} - (V_a + V_b)$$

A merger adds value only if the two companies are worth more together than apart, that is, when synergy is positive. Synergy is attributed to higher sales, lower

Table 3 M&A versus internal development

	Mergers and acquisitions	Internal development
Lower cost	+	
Acquire know-how	+	
Quicker	+	
Higher risk		+
Acquire new markets	+	
Acquire new products	+	
Synergy	+	

Source: Harvey Poniachek

Table 4 Sources of synergy

Synergy is attributed to the expansion of output and revenue, or *a reduction of cost thanks to avoidance of duplications*, such as administrative costs and overhead costs *and/or* economies of scale

Increase in revenue is often related to expansion of the customer base and cross selling.

Financial synergies related to decreased costs of capital through lowered risks, better cash flows, and increased financial margins

Source: Damodaran (2001) and Bruner (2004)

cost thanks to removal of duplications in the post-merger and economies of scale and scope, lower taxes, and lower cost of capital. If the transaction is in terms of cash for stock, cost is defined as follows:

$$\text{Cost} = \text{Cash paid} - V_b$$

The net present value (NPV) of the transaction is as follows:

$$\text{NPV} + \text{Gain} - \text{Cost} = \Delta V_{ab} - (\text{Cash paid} - V_b)$$

There are various theories of mergers and acquisitions (Weston et al. 2004), but practitioners formulated models in terms of several variables. For instance, Bain & Company (2014), M&A model consists of five stages, as follows:

- M&A strategy – make you M&A an extension of your growth strategy, identify the target, and formulate how will M&A create value
- Deal thesis – examine how each deal creates value, apply existing capabilities to enhance value to the target
- Due diligence and valuation – test the deal rational, set a walk away price, determine where you can add value
- Merger integration planning – plan your post-merger integration and determine what is needed to integrate, and integrate where it makes sense
- Merger integration execution

McKinsey & Company (Goedhart et al. 2010) identify the strategic rationale for an acquisition that creates value as follows:

- Improving the performance of the target company
- Removing excess capacity from an industry
- Creating market access products
- Acquiring skills, or market technology more quickly or at a lower cost than they could build in-house
- Picking winners early and help them develop their businesses

Reasons for mergers include:

- The prospects for synergy where $2 + 2 = 5$ that arise from lower cost, higher revenues, and higher profit.
- It may be cheaper to acquire another company than to invest internally, or the target company may be undervalued and it represents a good investment.
- Diversification – It may be necessary to smooth-out earnings and achieve more consistent long-term growth and profitability.
- Short-term growth management may be under pressure to turnaround sluggish growth and profitability.
- Some mergers are executed for “financial” reasons and not strategic reasons.

The world’s most successful companies rely heavily on acquisitions to achieve their strategic goals faster, cheaper, and less risky than organic expansion (Notop 2007). For instance, Pitney Bowes achieved great success by developing and applying a merger process with more than 70 acquisitions that increased the company’s revenue by more than 25% and made a significant contribution to net income and cash flow. The following five rules had been critical in the acquisitions process that could be effective for meeting strategic objectives if executed properly.

- Rule 1, stick to adjacent fields of your business.
- Rule 2, manage acquisitions as you would a portfolio of investments.
- Rule 3, the leaders of business units are in the best position to gauge potential acquisitions and need to be the sources and owners of acquisition proposals, whereas the corporate development group’s role should be to facilitate and execute the details of the transaction.
- Rule 4, be clear on how acquisitions should be assessed in terms of various criteria, including expected synergy and what integration to expect.
- Rule 5, avoid impulsive purchasing.

The merger process could be described in terms of an input-output flow chart as shown in Table 5 and Fig. 3.

102.2.1.1 Takeover Tactics

Hostile takeovers have declined from 14% of total deals in the 1980s to 4% in the 1990s. The main tactics for hostile bidders are:

Table 5 The mergers process and decision-making

1. What to acquire? ID target
2. Where to acquire? Location decision
3. What price? Valuation process
4. Considerations (means of payments)
5. Expected benefits/synergies?
6. Environmental constraints: legal, tax, accounting, operational/competition
7. Proposed structure
8. Post-merger integration
9. Approach the target, convene preliminary discussions, propose the deal structure, and formulate a letter of intent
10. Conduct due diligence and verify the facts and data
11. Analysis – expected benefits/synergies? $1 + 1 = 3$ or $= 4$
12. Assessed benefit and price
13. Is it an attractive target?
14. Consideration (=means of payment): cash, stock swap, other securities, or a mix of the above
15. Environmental constraints
16. Legal (state corporation law, corporate bylaws, securities laws, fiduciary responsibilities)
17. Antitrust (DOJ and FTC)
18. Tax (tax deferred or taxable?)
19. Accounting (purchase methods – newly issued requirements)
20. Operational implications
21. Complete negotiations, sign the agreement, make the announcement, and close the deal
22. Begin or continue the merger integration
23. The mergers process is assisted by <i>investment bankers</i> , attorneys, accountants, tax specialists, and others
24. Some large firms, however, conduct most of the M&A process internally (e.g., GE, Intel, Cisco Systems)

Source: Harvey Poniachek

- A bear hug – Exert pressure on target's management before launching a hostile bid by contacting the BODs with an expression of interest in acquiring the target that could sometime follow by a public announcement on the intent to acquire the target or by mailing a summary of the offer to the Wall Street Journal.
- In a friendly transaction, the tender offer will be made pursuant to a negotiated merger agreement with the target company and its board of directors. A friendly merger that is funded by cash requires the acquirer to file a proxy statement with the SEC and issue to shareholders and to be approved at a shareholders meeting and closed. A friendly merger funded with stock requires the issuing of a proxy statement with the SEC and the securities offered in exchange for the target shares need to be registered with the SEC. Upon completion of the proxy statement and share registration, both documents are mailed to shareholder + a proxy card to be filled out by the shareholders. The deal is voted on the shareholders meeting.

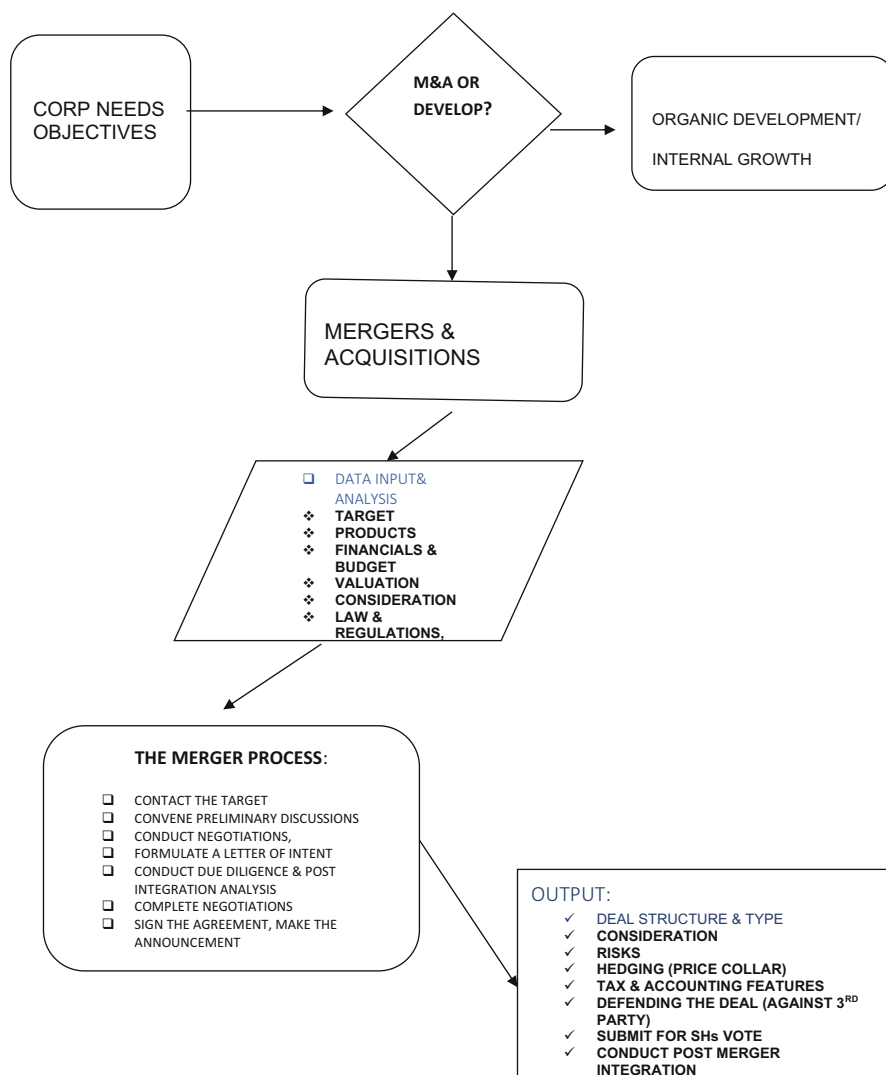


Fig. 3 The merger process. (Source: Harvey Ponichcek)

- A hostile bid is usually structured as a tender offer to the target shareholders. A hostile deal funded with cash requires the acquirer to issue a tender offer to the target shareholders consistent with the William Act requirement. Due to the short timetable, the SEC does not have the opportunity to review and comment on the materials, but it might do so subsequently. Hostile deal funded with stock for stock requires filing a registration statement with the SEC by the bidder prior to launching the tender offer. The statement needs to be cleared by the SEC prior to its dissemination to shareholders.

- A tender offer issued to the target's shareholders skips the BODs and management. The tender offer is in compliance with the Williams Act. The hostile tender offer is a two-step process: in the first step, the hostile bidder acquires control of the target by securing sufficient shares (in Delaware it is 85%), and in the second step, which is known as a back-end or close out transaction, the remaining share are purchased. If the buyer acquires a very large percent of the shares, it could freeze out or close out the remaining shareholders via a short-form merger.
- A proxy fight is an attempt by a single shareholder or a group of shareholders to take control or bring about changes in a company's conduct through a proxy mechanism of corporate voting. A bidder is attempting to use his voting rights and obtain support from other shareholders to oust the incumbent bod and/or management. The SEC regulates proxy contests and proxy solicitation, which are subject to section 14(a) and requires public disclosure within 10 days on schedule 14a. There has been a sharp increase in proxy fights that is partially explained by aggressive activity of hedge funds to enhance shareholders value. The two main types of proxy contests include contests for seats on the board of directors, and contest relating to management proposals.

102.2.2 M&A Transactional Forms

Mergers commonly commence with direct discussion and negotiations between the acquirer and the target company or through investment bank intermediation. If negotiations are successful, the acquirer and the target company enter into an agreement that sets the terms and conditions of the acquisition and the structure of the transaction. Companies involved in M&A negotiate confidentiality agreements or letters of intent that are referred to as a "Non-Disclosure Agreement" or "NDA." A letter of intent sometimes referred to as a "memorandum of understanding" or "MOU" is more common in private transactions than in public company deals. The transaction structure has important implications for the tax treatment of the transaction and the speed at which the transaction will be completed.

In a merger, two or more corporations are merged into one of the entities, which is the surviving entity of the merger. In a consolidation, the entities are consolidated to form a new entity and the stock of the non-surviving entity or entities are exchanged into stock of the surviving or resulting entity, cash, or some other consideration. Acquisition of a US public company could utilize state merger statutes, where the parties enter into a merger agreement that requires the approval of the board of directors and the shareholders of the target company (Hall 2014; Lathan and Watkins 2020).

There are several versions of statutory mergers, including

- In a statutory merger, the target company is merged into a newly formed subsidiary of the acquirer – through a forward triangular merger structure – and the surviving entity becomes a wholly owned subsidiary of acquirer. The target's stockholders receive stock, cash, debt, or a combination thereof from the acquirer.
- In a reverse triangular merger, a newly created subsidiary is formed by the acquirer that is merged into the target, and target's stockholders receive stock, cash, debt, or

a combination thereof, of acquirer in exchange of their shares. The target becomes wholly owned subsidiary of acquirer and its corporate identity is preserved.

- In an asset purchase, the acquirer may purchase either all or part of the assets of a corporate target in return various consideration, and joint venture.

The merger and acquisition of a US company can be made on a friendly basis, pursuant to negotiation with the target's board of directors, or through a hostile approach, without the involvement or approval of the target's board of directors. A friendly transaction affords due diligence on the target, cooperation of the target's management, faster closing and lower costs. A hostile transaction involves considerable complexity and risk, and it is suitable when negotiating fail, and most often hostile offers serve as a means to bring a target to the bargaining table, rather than as the ultimate means of acquiring a target company.

M&As are motivated by several factor, including:

1. Interest in to exploiting "synergies" that could reduce cost, expand output and efficiency.
2. Reduction of cost through economics of scale and scope.
3. To acquire work forces with particular skills.
4. Obtain know-how.
5. Benefit from lower taxes.
6. To diversify or expand markets.
7. *Acquire an undervalued at "bargain" price, and*
8. *Managers interest in "empire-building."*

In a one-step statutory merger, the acquirer negotiates a merger agreement with the target's board of directors, which then needs to be approved by a majority of the target's shareholders. The merger transaction is usually structured in the form of a "reverse triangular merger." If the consideration includes securities, those securities need to be registered with the SEC, and the acquirer needs to file a registration statement with the SEC.

In a two-step acquisition, an acquirer pursues a hostile takeover by issuing a tender offer (to acquire the shares of the target subject to the Williams Act) directly to the target shareholders, in exchange for the cash or the acquirer's securities. Following the purchase of the tendered shares, the acquirer and the target will complete a back-end statutory merger (most often via a reverse triangular merger structure). The vast majority of US public company are one-step mergers involving consideration consisting in whole or in part of stock.

102.3 M&A Financing

Two of the most common forms of financing for acquisitions are the use of debt or the issuance of equity. Issuing debt has tax benefits because the interest payments are tax deductible. Another benefit of issuing debt is that no additional shares are issued and so there is no dilution of ownership. However, issuance of excessive debt could

hurt the company's credit rating which would hinder its ability to borrow money in the future and would lead to an increase in the company's cost of debt and debt issuance may be limited by existing lender covenants that set restrictions on the amount of debt the firm can assume.

Despite the higher cost of equity, it is quite commonly used in M&A transactions because of the flexibility it provides the issuers. Some of the benefits of equity include (i) no mandatory interest payments, (ii) no principal that must be repaid, and (iii) no restrictive covenants related to its issuance. Financing an M&A transaction with equity has no impact on a company's credit rating therefore allowing the company to issue debt in the future if needed. However, equity issuing could hurt a firm's earnings per share and return on equity as it becomes less leveraged. Frequently, public companies use equity financing as their preferred form of payment in M&A transactions, but debt still plays an important role because of its cost effectiveness and the advantages of leverage.

In 1988, nearly 60% of all large deals (of over \$100 million) was paid for in cash and less than 2% was paid for in stock; in 1998, 50% of all large deals was paid for in stock (Rappaport and Sirower 1999); in 2010, the cash and stock were approximately 50/50 (Malmendier et al. 2015); whereas in 2016, 33% of all deals was paid in stock and 17% in cash. In a cash deal, the acquirer takes the entire risk that the acquisition will fail, while in stock transactions, risk is shared proportionately with the ownership of the target company between the two parties. Paying with stock preserves cash and avoids the need to borrow to fund the deal and provides the seller the opportunity to share in the future growth of the business and defer taxes on capital gains associated with the sale.

A stock-for-stock offer could be structured as a fixed number of shares or a fixed value of shares. In a fixed exchange ratio, share price between the announcement date and the closing date could change but the exchange ratio will stay fixed, thus the total deal value is not actually defined until closing. Sometimes the exchange ratio floats to ensure a fixed transaction value. Table 6 provides a numerical illustration.

Table 6 Cost of acquisition: cash versus stock

	Firm A	Firm B	Cash	Shares	Shares, amended
Market value (V_a , V_b)	2000	200	550	2530	2530
Number of shares	100	20	25	115	113.64
Price per share	20	10	22	22	22.263

Source: Harvey Poniachek, Adopted from Ross et al. 2019, p. 891

Notes

1. Firm B, the target, requested a 50% premium or \$300 and the value of both firms is assumed at $(2000 + 200) \times 1.15 = \2530 . Synergy equals $\$2530 / (2000 + 200) = 15\%$ or \$330
2. The acquisition NPV = Synergy - Cost = $\$330 - \$100 = \$230$
The value of Firm A share following the merger is $2530 / (100 + 15) = \$22$
3. Amended shares offer for Firm B = $2.530 \times R = 300$, $R = 0.12$
4. $0.12 = \text{New shares issued} / (100 + \text{New shares issued})$, New shares issued = 13.64
The price per share = $2530 / (100 + 13.64) = 22.263$

102.4 Restructuring and Divestitures

The primary restructuring approaches are conducted through the following functions:

- Sell offs or divestitures.
- Equity carve outs.
- Spin offs.
- Split offs.
- Split ups.
- Tracking stock (targeted stocks).
- Other strategies (JVS, ESOPS, LBOs), and
- Financial reorganization (debt-equity exchange offers, leverage recapitalizations, financial reorganization under Ch. 11).

A divestiture is the sale of a portion of the firm (assets, product line, sub) to an outside party for cash and/or securities or a mix of both. Some 40–50% of all M&As are divestitures by other firms. Equity carve outs involve the sale of an equity interest in a subsidiary to the public via an IPO. It begins with the creation of a new legal entity with its own management and board. Spin offs sometimes follow as the second stage of an equity carve out, in which some of the subs' shares are IPOed (less than 20%) and then spun off to shareholders of the parent on a pro rata basis. The parent often maintains 10–20% shareholding in the new entity.

Examples: In 1999, GM has done an equity carve out of Delphi via an IPO (17.7%), followed by a spin off (80.1%); GM kept 2.2%. The spun off entity becomes a separate legal entity, which is independently managed with its own management and board of directors. AT&T spun off NCR. In the basic spin off, controlled subsidiary's shares are distributed on a pro rata basis to the company's existing shareholders. Kraft Foods spun off Altria Group, Inc. planed on spinning off 88.5% of Kraft Food, Inc. (1/26/07, WSJ), Philip Morris Co (which became Altria) IPOed Kraft in 2001 by selling 280 million shares to the public. Altria shareholders were expected to receive 0.69 Kraft share for each share of Altria. Following the divestiture, Altria would become three distinct companies (1) Altria, (2) Phillip Morris Int'l, and (3) Kraft Foods.

Spin Offs Tax Effects. To qualify for a tax-free spinoff, the parent must own 80% of the sub's voting shares and 80% of the sub's shares need to be distributed. In cases where the spinoff is the second stage of an equity carve out, 80% of the shares in the second stage need to be distributed to qualify for a tax-free reorganization. Companies sometimes employ a two-stage spin off: the first stage IPOing by selling nonvoting shares, and in the second stage spinning off, the voting shares in tax-free transaction.

- Split offs.
- A new legal entity is created with its own management and board. Some of the shareholders exchange their shares for new shares in a division or subsidiary that is split off the parent company.

- There is a need to apply valuation and the exchange ratio to facilitate the exchange of shares.
- Example: In 1999, Dupont IPOed Conoco, which then followed by a split off. AT&T 1984 antitrust agreement with DOJ split up the corporation via spin offs into seven regional baby bells.
- Split ups.
- Split ups are usually accomplished by spin offs. In a split up, the entire firm is broken up through a series of carve outs and spin-offs with the parent ultimately ceasing to exist. Example: In 1999, Hewlett-Packard split into two firms, IPOed 15% of Agilent followed by spin off.
- AT&T restructured in 1995 by splitting into three public companies, where lucent was IPOed and then spun off.
- Divestiture/spin-off process.
- Decision to divest/spin-off relies on financial analysis (like an acquisition) by utilizing the DFC approach (WACC-NPV, APV ON FTE).
- Restructuring plan and agreement between the parent and subsidiary relating to assets, liabilities, equity, and operating relations require a proxy statement, voting, and approval by shareholders.
- Shares issued in an equity carve out or spin-off must be registered with the SEC.
- Tracking stock.
- First issued in 1984 when GM acquired EDS and issued GM class E shares. By 1992, there were four tracking shares, including GM E, GM H (Hughes Electronics), and two USX issues (Marathon – oil, US Steel group). By 2000, there were more than 44 issues.
- A company splits its operation into two or more entities or segments and issues tracking shares for one or several segments. The segments remain wholly owned by the parent company which itself might be publicly traded.
- Targeted stock is issued via an IPO and/or to existing shareholders afforded some rights in the segments relating to dividends, but no other rights are afforded in the parent firm. The benefits of targeted stocks: the capital market is provided with pure play and values the different businesses based on their own performance. Enhance flexibility to raise funds; provide opportunities in acquisitions; management incentives are more properly reflecting performance.

102.5 Leveraged Buyouts (LBOs)

LBOs are financial techniques that utilize substantial debt to acquire a target company by management or an outside company that are affected by the level of interest and availability of debt financing. These transactions increased dramatically in the 1980s, with a most notable LBO by KKR acquisition of RJR Nabisco \$32.5 billion in 1988. Sources of LBO gains are attributed to higher growth and enhanced efficiency. Asset-based lending of senior debt usually accounts for 50–60% of the deal, subordinated debt for 10–20% of the deal, and equity for 20–30%. High-yield bonds or junk bonds provided a major source for LBO funding.

In a leveraged buyout an acquirer takes control of a target firm by borrowing against its assets and/or cash flows. LBOs grew dramatically in the USA during the 1980s, declined in the 1990s, and reemerged in the 2000s fueled by cheap debt in the CDO markets (Renneboog et al. 2017), and it came to a halt with the demise of the securitized debt markets at the end of 2007. Tax advantages and the expropriation of non-equity stakeholders are the main sources of wealth gains from going private. Other potential sources of wealth gains are stronger performance and value, the reduction in wasting corporate resources, and improved monitoring capabilities embedded in the governance structure of an LBO.

In a leveraged buyout an investor acquires a target company using mainly borrowed funds. The debt is serviced either by the acquired target's operating cash flows, or from the sale of a portion of the company, or from the refinancing of the acquisition by collateralizing the target company. The debt used to finance an LBO ranges from 50% to 80% of the purchase price and may be as high as 95%. LBOs typically reduce a company's tax payments (and hence increase the returns available to its equity investors), primarily due to the tax shield from interest expense, increased depreciation and amortization expense following the write-up of a company's assets upon its LBO.

The leveraged buyout boom of the 1980s was spearheaded by KKR, The Carlyle Group, Blackstone, and Apollo (Shleifer et al. 1987) who created limited partnership (LP) that raised funding from large pension funds, insurance companies, financial institutions, and high net worth individuals, to finance the equity portion of the acquisitions. Management buyout transactions led by the existing management are more common in Europe and Asia than the USA. Drexel Burnham Lambert and its head of the high-yield bond business, Michael Milken developed the high-yield bond market⁴ which provided significant sources of funds for LBOs. Bank loans were another source of funds for LBOs, where lead banks commonly reduced their exposure by syndicating the loans to other banks. In the mid-2000s, a combination of lower interest rates, loosening lending standards, and creation of loan funds known as Collateral Loan Obligations (CLOs) – induced an LBO boom.

A successful LBO execution needs the cooperation of a private equity firm, a commercial or investment bank to provide the acquisition loans, and an investment bank that underwrites the bond debt. A shell company is typically created to serve as the vehicle for the acquisition that then bids for all the shares of the target company. If it is successful in acquiring a majority of the shares, it merges with the target company to create the surviving entity, which will be responsible for servicing the acquisition debt.

102.6 Joint Ventures and Strategic Alliances

Joint ventures (JVs) and strategic alliances could provide alternatives to M&As. In a joint venture, two or more companies combine certain assets and work toward achieving specified business objectives, while the companies maintain their own

business independent of the JV. JVs are subject to the antitrust laws as administered by the DOJ and FTC.

JVs are created for specific objectives and are of short duration. The objectives include conduct R&D, gain access to key supplies, enhance distribution systems, gain access to foreign market, and an interim step to enter an unfamiliar market. Strategic alliances are less formal associations between companies than JVs and could be defined as a collaboration by two or more firms to achieve specific goals.

A joint venture represents a combination of subsets contributed by two or more business entities for a specific business purpose and a limited duration. JV participants continue to exist as separate firms and the JP representing a newly created business enterprise. The JP can be organized as a partnership, a corporation, or any other form of business organization.

102.7 Anti-takeover Defenses

A corporation can adopt preventive anti-takeover defense measures against unknown threats or take reactive action in response to a potential or actual threat. Effective anti-takeover defenses enable management to resist unsolicited takeover bids, protect the corporation's long-term policy, search for a better offer to maximize shareholders wealth, and stand against shareholders who challenge their decisions to merge (Klausner 2013). The common preventive takeover defenses that are applied by most Fortune 500 companies include (Gaughan 2015):

1. The shareholder rights plans, also known as poison pill
2. Corporate charter amendments

Following the global financial crisis of 2007–2008, an increasing number of US public companies adopted or renewed their shareholder rights plans after years of decline in the use of this defense mechanism (Klausner 2003). In the post-financial crisis, unsolicited and hostile transactions constituted over 22% of all deal volume in the USA, compared with 12.1% for 2007 (Reuters 2008). In 2008, a total of 127 companies adopted rights plans and 182 companies allowed their existing rights plans to expire.

102.7.1 Anti-takeover Defense Measures

Pursuant to Delaware General Corporation Law (DGCL), the board of directors holds exclusive power to manage the corporation and has ultimate control of the corporation's business decisions, including business decisions associated with takeover activity. Shareholders have three basic rights as owners of the corporation that enable them to monitor managers (Thompson 1999) the right to sell their shares freely to third parties; the right to sue on behalf of the corporation, which can also be used as a tool to monitor managers; and the right to vote on significant corporate

matters that affect the corporation's destiny, including annually voting for the election of the board of directors. Thus, managers who do not perform in line with the expectations of their shareholders will be replaced by them, either by the shareholders tendering their shares to an acquirer who will replace the management or by the shareholders replacing the management at the annual meeting. The shareholders select the board of directors at the shareholders' annual meeting for a 1-year term, and if they are not satisfied with the board's performance or disagree with their decision-making, they can replace the entire board, a majority of the board, or those directors which they wish to replace at the annual meeting.

102.7.2 Shareholder's Rights Plan or the Poison Pill

The shareholders' rights plan, also known as the poison pill, is the most effective defense measure against a hostile takeover.¹ Since its inception, the poison pill was triggered in a very small number of occasions (The Delaware Chancery Court 2009). The strong anti-takeover effect of the pill either forced bidders to negotiate a friendly deal with the board or took corporations out of the takeover market. In 2011, the poison pill was challenged in the Delaware Supreme Court's *Airgas Poison Pill* case, but the court upheld its legitimacy.

Hostile bidders can go directly to the target shareholders with a tender offer to acquire their shares. The poison pill disables this possibility by making the takeover too expensive and economically unviable. This feature of the defense measure forces the hostile bidder to negotiate with the board a "friendly" deal (Lipton 1987).

The poison pill can be structured as a (i) the flip-in right or (ii) the flip-over right. Under the flip in, the board issues a right in the form of a special dividend to shareholders for each of their common share that entitles them to purchase common stock in the target at a 50% discount when a hostile acquirer (or group of acquirers) crosses a certain ownership threshold. Following the exercise of the rights, all shareholders increase their holdings in the target, except for the hostile bidder whose rights are void (Lipton and Steinberger 2015). The excluding of the hostile bidder from the shareholders rights plan results in a substantial dilution in the shareholdings of the hostile bidder. The flip-over right is intended to protect shareholders against a squeeze out that might follow by a merger. The triggering event for the exercise of the flip-over right is the acquisition of 50% or more of the corporation's assets. Upon the triggering event, the shareholders receive a right to buy shares in the surviving corporation or the acquirer. The board has the exclusive power to install a poison pill and redeem it. If the board decides to redeem the pill, the

¹The poison pill was invented in the 1980s by Martin Lipton, one of the founding partners of Wachtell, Lipton, Rosen & Katz, a distinguished Corporate and M&A law firm based in New York. In the mid-1980s, poison pills were challenged at the Delaware Supreme Court and their validity was upheld.

Table 7 The effect of a poison pill on a hostile takeover party

<i>Pre-poison pill</i>	
Target's shares outstanding	100,000
Price per share	\$10
Hostile party acquires % of total shares	10%
Hostile party's shares	10,000
Cost	\$ 100,000
Outstanding shares held by others	90,000
Hostile party seeks to acquire 50%	50,000
Hostile party needs additional shares	40,000
Total hostile acquisition cost	\$ 500,000
<i>Post-poison pill effect</i>	
Double outstanding shares	180,000
Outstanding shares	190,000
Doubles 90,000 + 10,000 held by hostile party	
% of hostile shares after PP	5.3%
Post-PP cost for hostile to reach 50% control	\$ 850,000
Initial cost pre-PP	\$100,000
Total cost to hostile	\$950,000
Post-PP cost/pre-PP cost	190%

Source: Harvey Poniachek

Notes: PP = poison pill

Price per share is assumed to stay unchanged under the various scenarios

If the hostile party does not obtain the rights to acquire shares at a discount, the acquisition cost is approximately 190% of the cost without the pill. However, if the hostile party gets the pill, the cost with and without the pill is the same

redemption is made for a nominal amount. See Table 7 on the effect of the pill on a hostile takeover party.

Bidders could avoid the poison pill by negotiating a friendly deal with the board or launch a proxy contest to replace the board with members that will redeem the poison pill. A poison pill can be easily adopted by the board, harder for the shareholders to dissemble, and more difficult for a bidder to overcome. There is a decrease in the number of rights plans due to rising influence from activist shareholder who oppose the poison pills because it makes a hostile takeover almost impossible. The Institutional Shareholder Services (ISS), a proxy advisory firms that advises on shareholders voting matters, typically support pro-takeover corporate governance. Their guidelines usually recommend either the redemption of poison pills or adoption of a short-term poison pill, the renewal of which will be subject to shareholders approval.

102.7.3 Charter and Bylaws Provisions

A vast majority of corporations with IPO's include staggered board provisions in their certificate of incorporation at the time of the IPO. A staggered board makes it

more difficult for a hostile takeover party to control the board because a shareholder or a group of shareholders would have to win two consecutive elections to replace a majority of the board (Coffee and Palia 2014). Anti-takeover defenses are usually perceived as impairing shareholder power and increasing the board's governing power in the corporation.

The poison pill and staggered board are widely recognized as the most effective anti-takeover defense measures, but there are other measures that a corporation can adopt as part of its defense profile. A provision in the corporation organizational documents that prevents shareholders from packing the board. Pursuant to the DGCL, all matters other than election of directors require majority of the shareholder vote. A supermajority vote requirement, typically 75% of the shares entitled to vote, limits the rights of stockholders to amend certain charter and bylaws provisions to facilitate takeovers. Certain business decisions that require shareholders vote, such as a sale of the corporation, would require a higher percentage of shares to approve, making it more difficult for the hostile bidder to receive the approval of the shareholders and complete the takeover.

102.7.4 Reaction (Economic Defense)

Another type of defense measure is financial defenses that would usually be used once an actual takeover activity is pursued or if the corporation feels vulnerable and become subject to a hostile activity.

102.7.5 Pac-Man

Pursuant to this tactic the target becomes an acquirer. Instead of defending against a hostile takeover, a target that has enough resources may seek to acquire the hostile bidder to eliminate the takeover threat.

Pac-Man tactics are not widely spread nowadays, although we did recently see an example of a successful Pac-Man tactic by Men's Wearhouse against Jos. A. Bank, two strong men's apparel companies. The takeover actions between the two began when Jos A. Bank launched its initial bid to acquire Men's Wearhouse. The bid was rejected by the latter and eventually Men's Wearhouse gave Jos. A. Bank's shareholders an offer that they accepted. The company that was first a target saw the opportunity in the combination transaction and became an acquirer (Solomon 2013).

102.7.6 Golden Parachutes

Golden parachutes are rewards set in employment agreements of senior executives to compensate them, among other things, upon the occurrence of a change of control event. Unlike most defense measures that we have included in this part of the chapter, a golden parachute would usually be implemented in the executive's

employment agreement when such executive joins the corporations which could be before any takeover activity was being taken against the corporation. Although this defense measure is a pre-action defense measure and not a re-action defense measure, we have included it under this part of the chapter because of its financial feature and anti-takeover effect resemblance to the remaining re-active defense measures mentioned here. The main goal of the golden parachute is to make the target less attractive to a takeover by forcing the hostile bidder to pay large amounts from the acquired corporation treasury. Golden parachutes can be as much as tens of millions of dollars once a triggering event, such as a change of control occurs. A golden parachute might not be as an effective defense measure against such tactics as it would be against a change of control event. As of 2015, 70% of fortune 500 companies offered golden parachutes in the event of a change of control to select executive officers (Nelson and Watson 2015).

102.7.7 White Knight and White Squire

Targets that face actual takeover activity by a hostile bidder can recruit a second bidder, a white knight, to acquire the corporation in a friendly deal that will be negotiated with the management instead of being taken over by the hostile bidder. For a white knight to be successful and complete the acquisition, such white knight must outbid the hostile bidder and provide the shareholders a higher value.

102.7.8 Limitations on the Adoption of Anti-takeover Defense Measures

When a corporation faces a takeover threat, the board could adopt anti-takeover defenses if (i) there is an existing threat to corporate policy, and (ii) the defense measure adopted is “reasonable in relation to the threat posed.” When it is clear that a change of control will occur, the sole job of the board is to sell the corporation at the highest price available for shareholders and they are not allowed to take defense actions to stay independent and fend off bidders.

102.8 Deal Protection and Deal Certainty²

Merger agreements include various deal protection to ensuring that the transacting parties remain obligated to consummate the transaction. The key provisions are deal protection devices such as break-up fees, no-shop clauses, force-the-vote provisions, and shareholder voting agreements.

²This section follows Wachtell, Lipton, Rosen & Katz, Takeover Law and Practices, March 2016, <http://www.wlrk.com/files/2016/TakeoverLawandPracticeGuide.pdf>

A break-up fees payable by the target if it terminates the merger agreement to accept a superior proposal. Termination fees of approximately 2–5% of the merger value may deter third parties from acquiring the target that has reached a merger agreement. Break-up fees can be triggered if target's shareholders fail to approve the merger.

A “no-shop” provision in a merger agreement provides that a selling company will not encourage, seek, solicit, provide information to or negotiate with third-party bidders, but the seller is allowed to respond to unsolicited offers by supplying confidential information and to consider and negotiate with respect to certain competing bids. “Go-shop” provisions allow the target company to actively solicit competing offers after the merger agreement signing – usually 30–50 days.

Public company merger agreements generally include provisions requiring the board of directors of the target to recommend that shareholders vote in favor of the merger agreement, except in specified circumstances. Merger agreements often include provisions that permit a party to terminate the agreement to accept a superior proposal, subject to payment of a termination fee and other conditions – commonly known as a “fiduciary out.” A Delaware corporation may submit the merger agreement to shareholders vote even if the board subsequently changes its recommendation.

The “crown-jewel” lock-up is a device in which the target company grants the acquirer an option to purchase a key target asset if the proposed merger does not close. This device may serve to deter competing bidders, since even with a superior bid price, the competing bidders may not get the deal they are seeking (*i.e.*, at best they may get a deal without the crown jewels).

Virtually all domestic public company merger agreements allow the buyer to refuse to close if there has been a “material adverse effect” on or a “material adverse change” in the target company's business.

Traditionally, strategic buyers, with significant balance sheets, were expected to fully commit to the completion of a cash acquisition, whereas financial sponsors are allowed to exit the deal in the event that they were unable to obtain financing.

The reverse termination fee required the buyer to pay a fee in the event the buyer failed to close due to an inability to obtain financing (later expanded to a failure to close for any reason).

The consideration used in a particular transaction depends on the characteristics of the deal and the relative bargaining strength of the parties.

All-cash transactions have the benefit of being of certain, but require cash on hand or ability to borrow, and it has a tax effect on the seller.

All-stock transactions are subject to price risks between signing and closing. Such risk can be hedged by using a price collar.

102.9 The Board's Role in M&A

Corporate governance in the 1980s was dominated by intense merger activity distinguished by the prevalence of leveraged buyouts (LBOs) and hostility. In the early 1990s, substantial merger activity resumed in the second half of the decade,

while LBOs and hostility did not. Internal corporate governance mechanisms appear to have played a larger role in the 1990s (Holmström and Kaplan 2001). Takeovers challenged in courts produced judicial decisions relating to the duties of directors of the board, anti-takeover defenses, and concerns about directors potential conflict of interest. Most takeover cases were adjudicated by the Delaware courts where most of the fortune 500 are incorporated.

Most jurisdiction describe the directors as having two fiduciary duties: (1) The duty of care and (2) The duty of loyalty (Monks and Minow 2011). Duty of care requires a director to exercise due diligence in making decisions by relying on full information. Duty of loyalty requires a director to demonstrate unyielding loyalty to the company's shareholders. Additional and more specific definitions of the board's duties – beyond the legal definitions cited above – were suggested by the business roundtable and the American Law Institute.

The presumed dual fiduciary duties of directors constitute the business judgment rule or standard. The business judgment rule is the standard by which directors are judged when they exercise their fiduciary duties to shareholders during an attempted takeover (relating to pricing and process). The court will not second guess a director's business judgement decision if it can be demonstrated that he/she exercised all possible care and acted with due loyalty (*note that the burden of proof shifted to the director*). *In the M&A context, some special applications of the business judgment rule have developed through litigations challenging directors' conduct [see White & Cooperman].*

102.9.1 Seminal Court Cases on Corporate Governance

In the Trans Union case (1985), the court ruled that directors had not met their duty as fiduciaries by agreeing to sale the company without a sufficient verification process to obtain the best terms. The primary implication of the trans union case was that the decision-making process and not the substance of the decision was of importance, and the court did not rule on the business judgment. Following the trans union decision, the Delaware legislature amended the Delaware general corporation law relating to duties of directors in mergers/acquisition cases, as follows: *1. Once it is recognized that a company is "in play" and it is likely to be sold, directors' obligation is to obtain the best price for shareholders 2. Directors should be fair to all bidders; and 3. Self-interest should not impair directors' decisions* (Weston et al. 2004).

In Unocal V. Mesa petroleum (1985), Unocal anti-takeover defenses included a self-tender offer in which the target made a tender offer for itself in competition with the mesa offer. The Delaware supreme court held that the directors may have acted in their own self-interest in which they favored their own tender instead of objectively searching for the best deal for shareholders. In such instances, the directors are subject to enhanced scrutiny and must demonstrate that (1) based on their business

judgment they had reasonable grounds to believe that there was a danger to the pursuit of a corporate policy that was in the best interest of shareholders and (2) that their defenses were reasonable in relation to the threat.

In the Revlon case (1985), the court held that when it becomes clear that the sale or break up of a company is inevitable, the board has the fiduciary duty to maximize shareholders wealth – by shifting their focus away from defending the company to selling it at the highest price (these duties became known as the Revlon duties). The court established that rather than promote the auction process to maximize shareholders value, Revlon's anti-takeover measures inhibited it.

In the Revlon and Unocal cases, the courts addressed whether under certain circumstances there was a limit to the anti-takeover defenses that a board could undertake to rebuff an unsolicited offer to acquire it. In the *Paramount v. Time* (1989) case, the court stated that Time-Warner was a strategic merger not involving a change in control, therefore, the Revlon duties were not triggered, and the normal business judgment rule standard applied. In the case of *Paramount v. QVC Network* (1993), the supreme court of Delaware held that the directors of Paramount violated their fiduciary duties in not maximizing the selling price of their company. The bidding contest proceeded with Viacom ultimately acquiring Paramount.

Boards must embrace the principles of due diligence, buy versus build analysis, financial analysis, and risk assessment as primary factors in their decision-making and evaluation process for M&A as well as other capital investment proposals. M&A transactions should help drive shareholder value and be aligned with both short-term and long-term strategic objectives and be accretive to the market price of the company's shares as well as the brand, reputation, and consumer perception of its core products and services. Risk needs to be avoided and strategic objectives met.

The board's role in M&As is to oversee the company's management as it negotiates the contours of a particular deal; test and confirm the economic premise of the proposed transaction; confirm the value proposition that the deal offers to the company's shareholders; confirm that management has taken necessary steps to mitigate the risk of the transaction; and confirm that management is planning the post-closing integration adequately. Boards and company executives must address issues of valuation, deal structure tax and securities laws, and often face a buy or sell decision and other structural business decisions and spend about a third of their time with such matters.

The board plays a key role in overseeing the M&A deal and make sure that valuations are assessed properly, that due diligence is complete and that the post-merger integration is successful. The engaged board will develop the company's overall strategic direction and might determine that growth should be driven through acquisition. Successful mergers and acquisitions transaction must reflect the economic needs of both buyer and seller and must convey real and durable value to the shareholders of both companies. Achieving it involves a review and analysis of financial statements, a genuine understanding of how the proposed transaction meets the economic objectives of each party, and a recognition of the tax, accounting, and legal implications of the deal.

102.10 The Legal and Regulatory Framework of M&As

The US has a dual legal system of state and federal governments. Merges are governed by corporation laws of the states in which the companies are incorporated, but the federal government regulates transactions involving securities, antitrust policies, and business dealings in several regulated industries.

The US federal securities laws include the Exchange Act of 1933, the Securities Exchange Act of 1934, and the regulations under the Securities Exchange Commission (“SEC”). Tender offers are subject to the Williams Act of the Securities Exchange Act. In US mergers involving securities, the issuer must disclose important financial information through the registration of securities with the SEC. Each of the 50 states regulates securities transactions pursuant to Blue Sky Laws and almost every state requires that securities sold in the state be registered.

Antitrust issues are administered by the Federal Trade Commission (“FTC”) and the Department of Justice (“DOJ”) and, to a lesser extent, by states. The US antitrust law prohibits mergers and acquisitions that could reduce competition or create a monopoly. The Hart-Scott-Rodino Act (“HSR”) requires pre-mergers and acquisitions notification to the FTC and the DOJ by merging companies to enable the government to analyze the effects of proposed transactions on competition before the acquisitions are approved and consummated. The Committee on Foreign Investment in the United States (“CFIUS”) reviews transactions that could result in foreign control of a US business to determine the effect of such transactions on the national security of the US.

The US antitrust laws were first legislated in 1890 with the passage of the Sherman Act and then in 1914, the enactment of the Clayton Act and various refinements through amendment over time (Abbott 2005). The Hart-Scott-Rodino Antitrust Improvements Act (“HSR” or “HSR Act”) of 1976 requires companies to file pre-merger notifications with the Federal Trade Commission and/or the Antitrust Division of the Department of Justice. After filing the HSR form, the parties cannot close the acquisition for a 30-day waiting period. The government (or another plaintiff) may challenge a merger before it is consummated and seek an injunction blocking the merger, or it may institute litigation based upon alleged violations of the Federal antitrust laws. State attorney generals may also bring actions under the federal antitrust laws.

The Antitrust Division of the DOJ and the FTC have jointly issued Horizontal Merger Guidelines (“Guidelines”) (U.S. Department of Justice and the Federal Trade Commission 2010) outline the principal analytical techniques, practices, and the enforcement policy of the Department of Justice and the Federal Trade Commission with respect to mergers and acquisitions involving actual or potential competitors under the federal antitrust laws.

The Guidelines’ first step in the merger analysis is to identify the relevant product and geographic markets and assess whether the merger would significantly increase concentration in any of those relevant markets. Market concentration is calculated according to the Herfindahl-Hirschman Index (“HHI”) by summing the sum of

squares of the individual firms' market shares. The Agencies consider whether the prospect of entry into the relevant market through merger or acquisition could alleviate adverse competitive effects by expanding output and keep low prices to benefit customers. A merger is not likely to enhance market power if entry into the market could not raise price or otherwise reduce competition compared to the level that would prevail in the absence of the merger. The Agencies credit efficiencies that are likely to be accomplished with the proposed merger. A merger is not likely to enhance market power if imminent failure of one of the merging firms would cause it to exit the relevant market.

Various other federal agencies, including the Federal Communications Commission ("FCC") and the Federal Reserve Board ("FRB") apply additional regulatory requirements relating to mergers in media and communications, banking, utilities, insurance, and airlines. The Sarbanes-Oxley Act that was enacted into law in 2002 and mandated several reforms to enhance corporate responsibility, enhance financial disclosures, and combat corporate and accounting fraud. It also created the Public Company Accounting Oversight Board ("PCAOB") to oversee the activities of the auditing profession. The Dodd-Frank Wall Street Reform and Consumer Protection Act was adopted in 2010.

102.10.1 State Anti-takeover Laws

Delaware General Corporation Law ("DGCL") prohibits an acquirer of 15% or more of a company's outstanding stock from engaging for a 3-year period following the acquisition, in any business combination with the company. Over a third of the states authorize defensive action by a target's board to defend against a hostile bid, including adopting a shareholders' rights plan (a poison pill) without shareholder approval.

102.10.2 Tender Offer

The Securities Exchange Act requires disclosure of important information by anyone seeking to acquire more than 5% of a company's securities by direct purchase or tender offer within 10 days of the acquisition. Within 10 business days of commencement of a tender offer, the target must reply whether to accept or reject the offer, or express no opinion, or explaining that it cannot take a position at that time. If the acquirer completes the tender offer and obtains 90% of the target common shares (in some states, this percentage is as low as 85%), the acquirer can squeeze out the minority shareholders through a short-form merger, without the approval of any other shareholder of the target company, and acquire all remaining target common shares.

102.11 Cross-Border Mergers and Acquisitions

Cross-border M&As have become very popular and in many instances have emerged as a two-way street despite their risk and complexities. In essence, domestic and cross-border transactions are quite similar in terms of the process and analytical requirements. Multinational companies (“MNCs”) increasingly utilize cross-border mergers and acquisitions to avail themselves of significant opportunities to grow and enhance corporate competitiveness, diversify geographically, support their customers abroad, and gain access to natural resources. The European Union and the US were the largest acquirer and target countries accounting in 2019 for 72% and 85%.

The past three decades have witnessed a dramatic expansion in cross-border acquisitions, with the volume expanding from 23% of total merger volume in 1998 to 33% in 2015 (Erel et al. 2012). As the world’s economy becomes increasingly integrated, cross-border mergers are likely to become more important. In recent years, there has been a rapid growth of cross-border M&As flows to emerging countries, including China, India, and Russia, that now account for approximately one-third of overall activities (Bertrand and Betschinger 2011). Cross-border take-overs have become increasingly important, and the outlook is for continuous growth in the future (Humphery-Jenner et al. 2012; Morgan 2017).

The motivation and process of cross-border M&As are in the main like those of domestic M&A but face different legal regimes and economic circumstances (Rossi and Volpin 2002). The theories of cross-border M&As state that (Coeurdacier et al. 2009) gains arise from increased economies of scale or scope. Multinational capital budgeting (Eiteman et al. 2016) for investment in foreign countries applies the same theoretical framework as domestic capital budgeting with very few differences. In international investment, the issue is how to adjust for the various risks, either via the discount rate or the cash flows, but not both at the same time. Although adjusting the discount rate might be easier, adjustment of the cash flow is probably more appropriate.

Mergers and acquisitions are complex transactions with many potential problems and a high failure rate and poor performance due to poor management during the merger process. A typical M&A process commonly follows the following steps (Downey and Technical Information Service 2008):

- Strategic planning.
- Identify and evaluate of potential targets.
- Initial contact.
- Valuation of the target company.
- Due diligence.
- Deal execution includes the negotiation of the deal, and
- Integration – after closing is complete, the acquiring company must start the process of integrating the acquired company into their business.

A KPMG survey (KPMG 1999) found that only 17% of deals added value to the combined company, 30% produced no discernible difference, and 53% destroyed

value. Bruner reviewed the evidence from some 115 M&A studies from 1971 to 2001 (Bruner 2002) and observed that the conventional view that failure is the average outcome of all M&A is false. A company is *more likely* to fail in its M&A under the following circumstances:

- *It enters a fundamentally unprofitable industry.*
- *It lacks the necessary industry know how.*
- *The expected economic benefits (synergy) of the deal are improbable or not incremental to the deal.*
- *It lacks a competition advantage in the business.*
- *It is not very creative in deal structuring, and*
- *Poor post-merger integration.*

To value foreign companies for acquisition, we could apply the same methodologies that are used in valuation of domestic firms, but several international complexities need to be recognized that might affect the outcome (Koller et al. 2010), that include international taxation, forecasting cash flows in foreign and domestic currency, estimating the cost of capital in foreign currency, and incorporating foreign-currency risk in valuations.

A conference on valuation in emerging markets (Bruner et al. 2002) that included some 30 countries concluded that, first, there is currently no clear single “best practice” for the valuation of assets and securities in emerging markets; second, emerging markets differ from developed markets in areas such as accounting transparency, liquidity, corruption, volatility, governance, taxes, and transaction costs – that are likely to affect firm valuation.

Emerging market risks can be incorporated through DCF scenarios by developing alternative scenarios for future cash flows and discounting rates and then weighting the values by the scenario probabilities. The cost of equity for emerging markets is estimated by utilizing the Capital Assets Pricing Model (“CAPM”) and a global risk-free rate, market risk premium, and beta. The cost of debt is reflected by the local yield on corporate bonds to which we add a country risk premium. The Weighted Average Cost of Capital (“WACC”) is then derived for the subject company based on the relative shares of equity and debt. Since companies in emerging markets are often more volatile than in developed markets, we recommend assessing the DCF valuation results with a country risk premium and a multiples.

102.12 Valuation for Mergers and Acquisitions

The three major valuation approaches that are used in mergers and acquisitions (“M&As”) include the income approach, the market approach, and the asset approach, with the first two being the primary valuation methodologies. The main features of the valuation approaches (Arzac 2008; Koller et al. 2010; Damodaran 2006, 2012; Rosenbaum and Pearl 2013), shown in Fig. 4, are as follows:

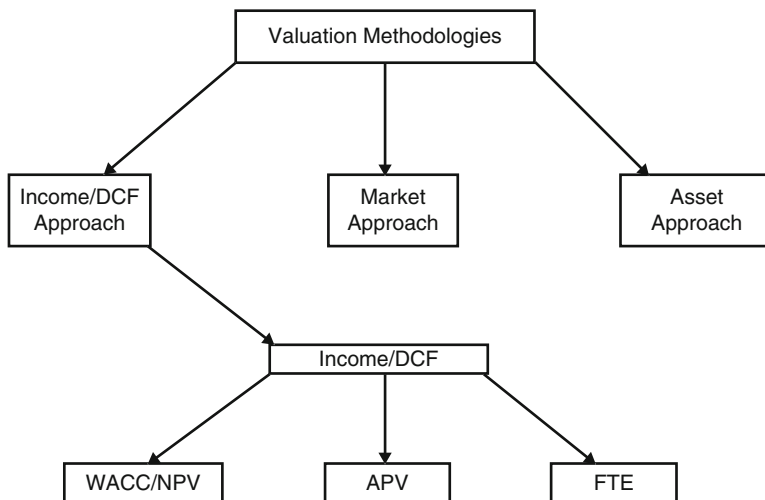


Fig. 4 Valuation methodologies for M&A. DCF = Discount cash flow, WACC = Weighted average cost of capital, NPV = Net present value, APV = Adjusted present value, FTE = Flow to equity. (Source: Harvey Poniachek)

1. The income approach could be applied in three different versions, depending on the nature of the cash flows, the capital structure of the firm or entity, and the discount rate. The approach derives the enterprise value (“EV”) of a target company based on the present value of the free cash flows that the company expects to generate in the future. The discount rate could be the Weighted Average Cost of Capital (“WACC”) or the cost of debt. The income approach requires a good measure of judgment due to the inherently uncertain business forecasts surrounding merger synergies (Brotherson et al. 2014).
2. The market approach assesses the enterprise value or equity based on comparable companies or comparable M&A transactions, by utilizing various multiples, such as the ratio of the price per share to earnings per share (“P/E”), or the ratio of the enterprise value to earnings before interest and tax (“EV/EBIT”), and several other multiples.
3. The net asset approach defines the value of common equity of a target company at book value as the difference between assets and liabilities, or by adjusting the assets and liabilities to market value and deriving their difference.

102.12.1 The Income Approach

The income valuation approach applies the Fair Market Value (“FMV”) or the investment value standard (King 2008; Fishman et al. 2007), and premises that the target company or entity is a going concern. The valuation determines the target’s worth, but the price paid for it would ultimately be negotiated between the buyer and

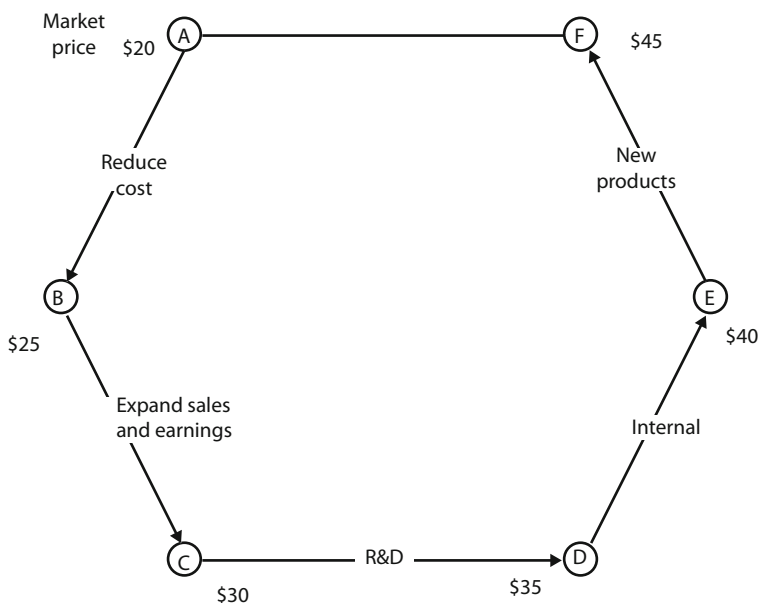


Fig. 5 The effect of synergy on the bidding price. (Source: Harvey Ponichchek). Notes: The target's market price per share is \$20, but different prospective acquirers can afford paying much higher prices for the target – depending on what they expect to do with the target following the acquisition. For instance, one prospective acquirer could obtain significant synergies via reduction of cost, expanding of sales and earnings, R&D, and international expansion into new markets. However, not every prospective buyer might have the same opportunities. Therefore, different prospective acquirers are likely to bid for the target at different prices

seller that would consist of the stand-alone value of the target company plus a fraction of the expected synergies. Because different prospective parties expect different synergies from M&As, their offer prices could vary considerably. See Fig. 5 on the effect of synergy and other influences on the bidding price per share.

The income approach applies the discount cash flow (“DCF”) methodology based on solid corporate finance theories (Berk and DeMarzo 2014; Ross et al. 2019) and could be applied to three different circumstances, as follows:

- First, the net present value (“NPV”) methodology, which is often referred to as the WACC methodology, discounts the unlevered free cash flows to the firm (“FCFF”) for the first 5 years and the terminal value to infinity. The terminal value is derived either through a formulary approach based on the last year's FCF, the growth rate to perpetuity and the WACC, or through a multiple of EBIT or EBITDA. The sum of the discounted free cash flow of the forecast period plus the present value of the terminal value determines the enterprise value – from which debt is deducted to derive the equity value. Data used in valuation is often measured with error, therefore, it is appropriate to use several valuation scenarios to estimate the target's value. We note that because the WACC was derived by

using the after-tax interest rate, this approach incorporates the tax benefit of debt implicitly through the cost of capital.

- Second, we could value an unlevered target company's free cash flows to the firm by discounting it at the unlevered cost of capital, and then add the present value of the interest tax shields attributed to debt discounted at the cost of debt, as well as other financial side effects, such as subsidies. This approach is called the adjusted present value ("APV") method.
- Third, the DCF method is based on free cash flows to equity ("FTE") discounted at the cost of equity for an unlevered entity or the cost of equity for a levered company.

Usually three valuation scenarios are derived, including high or optimistic, average, and low or pessimistic, and probabilities are assigned to each to obtain a probability weighted average scenario. The reasonableness of the valuations could be further examined through sensitivity analysis, including an application of a Monte Carlo simulation technique. Control and liquidity discounts and premiums could be applied as necessary (Pratt 2001).

102.12.1.1 Free Cash Flows

The free cash flows from assets equals to earnings before interest and taxes ("EBIT") plus depreciation, which is a noncash expense, minus taxes, minus capital expenditures ("CAPEX"), and minus changes in net working capital ("NWC"). Stated algebraically, FCF is defined as follows:

$FCF = EBIT + \text{Depreciation} - \text{Taxes} - \text{CAPEX} - \text{NWC}$, or alternatively stated as

$$FCF = EBIT (1 - t) + \text{Depreciation} - \text{CAPEX} - \Delta \text{NWC}$$

The definitions of the variables are as follows:

FCF = Free cash flows

EBIT = Earnings before interest and taxes

t is the marginal tax rate

$EBIT (1 - t)$ is EBIT after taxes

CAPEX = Capital expenditure, which equals $\text{Net Fixed Assets}_2 - (\text{Net Fixed Assets}_1 - \text{Depreciation})$

ΔNWC = Change in net working capital, which equals $\text{NWC}_2 - \text{NWC}_1$, and NWC equals $\text{Current Assets (CA)} - \text{Current Liabilities (CL)}$

The net working capital includes cash, receivables, inventory, and payables required for the operation of the business.

102.12.1.2 Terminal Value

A terminal value ("TV") reflects the value of all cash flows from the end of the projected period to perpetuity, and it constitutes a significant component of the

company's enterprise value, and often accounting for as much as 70% of the enterprise value. The terminal value could be estimated in two ways:

- First, through a formulary approach based on the last forecasted FCF, the growth rate, and the discount rate, as follows:

$$TV = FCF (1 + g) \div (WACC - g),$$

where

FCF (1 + g) = The expected growth of free cash flow for the year after the last year of the forecasted period, which is commonly year 6 and onward

WACC = The weighted average cost of capital

g = The expected constant annual growth rate of FCF (1 + g) in perpetuity

- Alternatively, the terminal value could be derived as a multiple of EBIT or EBITDA, where the multiple is derived from comparable companies or comparable M&A transactions. The differences between the two estimates of the TV could be attributed to differences in the discount rate, or growth rate or the expected risk.

102.12.1.3 The Discount Rate

The discount rate should reflect the risk associated with the free cash flows. Based on the capital structure of the company being valued, we derive a weighted average cost of capital or WACC, which serves as the discount rate for the free cash flows of the target company.

The formula for WACC is as follows:

$$WACC = (Wd) Rd(1 - t) + (We) Re,$$

where the variables are specified below:

Wd = The % share of debt in the capital structure, or $D/(D + E)$.

We = The % share of equity in the capital structure, or $E/(D + E)$.

D = Interest bearing debt.

E = Equity, or the market capitalization of the firm, which equals price per share x the number of outstanding shares.

Rd = Pretax cost of debt is the yield to maturity on corporate bonds with the same credit rating and maturity of the subject company or the cost of loans of the subject company.

T = Marginal tax rate.

Re is the cost of equity capital that can be obtained from the Capital Asset Pricing Model or CAPM, as follows:

$$R_e = R_f + \beta (R_m - R_f),$$

where

R_e = Cost of equity or the expected rate of return on the company's equity.

R_f = Risk-free rate of return on government securities over a time horizon consistent with the investment horizon. There are various risk-free rates that could be used, but the practice is to use Treasury Bills for up to 1-year maturity, Treasury Notes for up to 10-year maturity and Treasury Bonds for 30-year maturity.

β = The beta is a measure of the company's stock systematic risk or market risk or non-diversifiable risk, which reflects the company's volatility compared to the overall market index.

$(R_m - R_f)$ = Risk adjusted rate of return on the market index or the market risk premium, or equity risk premium. $(R_m - R_f)$ is usually estimated as the difference between the returns on the broad common stocks market index, such as the S&P500, and medium or long-term government bonds, depending on the horizon of the valuation. The market-risk premium ranges between 4.5% and 5.5% (Goetzmann and Ibbotson 2006; Fernandez et al. 2016).

R_m = Return on the market index, usually represented by the rate of return on the S&P 500 index.

The cost of equity is an estimate of an investor's required rate of return on an investment that could be estimated based on the CAPM; however, application of the CAPM to divisions, business segments, or private companies is hampered by lack of beta information. In such instances, we could derive the betas by using comparable companies through a process of unlevering and relevering, or by applying the Dividend Growth Model ("DGM") as an alternative to CAPM.

102.12.2 The Adjusted Present Value ("APV")

The adjusted present value ("APV") is designed to value operations or assets-in-place (Luehrman 1997; Ross et al. 2019), and values the business in a two-step approach by first valuing the operation, and then value the financial side effects and add both values together to constitute the value of the enterprise.

Specifically, we can value an unlevered target company's free cash flows by discounting it at the unlevered cost of capital, and then estimate and add the present value of the interest tax shields attributed to debt discounted at the cost of debt, as well as other financial side effects, such as subsidies. Because interest payments are tax deductible expenses, debt financing creates an interest tax shield for the firm, that is included explicitly or implicitly in the valuation of the firm.

In applying the APV, we first derive the pro forma income statement and balance sheet for 5 years by utilizing the percent of sales approach. Second, we derive the expected growth rate to perpetuity and the discount rate and derive the terminal

value. The free cash and the terminal value are present valued at the cost of equity for an unlevered company and form the value of the operating business.

In the second step of the APV, we value the financial side effects, which could include the debt shielding tax effect of interest, subsidies, cost of financial distress, hedges, and cost of issuing new securities. For instance, the financial side effects of interest tax shields equal $D \cdot r \cdot t$, where D is the interest-bearing debt, r is the cost of debt, and its terminal value is derived through the formulary approach similarly as was used in the WACC model. The present value of the tax shielding effect for 5 years and to perpetuity provides the value of the financial side effects.

The enterprise value of the firm is the sum of the value of operation plus the value of the financial side effects, that is, $APV = NPV + NPVF$, where NPV is the value of the unlevered firm and $NPVF$ is the present value of the financial side effects that are derived as follows:

- Determine the entity's value without leverage V_u , by discounting its free cash flows at the unlevered cost of capital, R_u , and with a constant debt/equity ratio, R_u is estimated by the following equation: $R_{wacc, u} = E/(D + E) R_e + (D/E) R_d$.
- Determine the present value of the interest tax shield in the following manner: Given the expected debt, D_t at time t , the interest tax shield at $t + 1$ is $D_t \cdot R_t \cdot t$.
- Add the unlevered value, V_u , to the present value of the interest tax shield to determine the value of the entity or company with leverage, V_L .

102.12.2.1 Flow to Equity ("FTE")

Determine the free cash flow to equity ("FCFE") through the following equation:

$$FCFE = FCF - (1 - t) * (\text{Interest Payments}) + \text{Net Borrowing}.$$

(FCFE) = Payment of dividends + repurchase of common equity, after taxes, debt repayments, new debt and preferred stock issues, and all reinvestment requirements.

$$FCFE = (\text{Net Income} + \text{Depreciation} - \Delta \text{Net Working Capital}) - \text{Gross Capital Expenditures} \\ + (\text{New Preferred Equity Issues} - \text{Preferred Dividends} + \text{New Debt Issues} - \text{Principal Repayments})$$

- Determine the equity cost of capital, R_e .
- Compute the contribution to equity value, E , by discounting the free cash flow to equity using the equity cost of capital.
- Discount the cash flow from the firm to the shareholders of the levered firm at the cost of levered equity capital, R_s , by the following three steps approach: (1) Calculate the levered cash flows, LCFs, (2) Calculate R_s .
- Value the levered cash flow at R_s .

The Market Approach

The market approach determines the value of a company or entity by comparing it to comparable companies or comparable transactions. The method is based on the derivation of various multiples, with the most common multiples are as follows:

- The P/E ratio of the price per share to earnings per share.
- The EV/EBIT ratio of the enterprise value to operating income.
- The EV/EBITDA ratio of the enterprise value to gross operating profit + depreciation and amortization.
- EV/FCF the ratio of the enterprise value to operating free cash flow, and
- EV/Sales the ratio of the enterprise value to company sales.

The financial multiples are selected from a sample of comparable companies or comparable M&A transactions and applied to the target company to estimate its range of values. Ideal comparables, also referred to as guideline companies, are in the same industry as the subject company being valued, and same markets, products, growth, profitability, and business risk. The market approach provides a popular valuation methodology, but the reliability of the outcome depends on the goodness of the comparables.

If comparable companies or guideline companies are used, a decision needs to be made whether to use Market Value of Invested Capital (MVIC) or Market Value of Equity (MVE). MVIC is defined as the entire invested capital value of the firm, including equity investment and debt investment.

There are a variety of sources for comparable or guideline companies, including the SEC 10-K filings on the SEC website or through EDGAR; S&P Compustat and S&P Capital IQ, which are searchable database of US and foreign public companies. Comparable transactions, or comparable deals, are commonly used alongside comparable companies. The difference between comparable companies and comparable transaction multiples is that the latter will reflect a “control premium,” typically 30–50%, that is not present the comparable companies. The “control premium” depends on the uniqueness of the assets and to what extent there are close substitutes for the technology, expertise, or capability in question, the distribution of financial resources between the bidder and target, or the possibility that the ex-ante target price was unduly inflated by market rumors. Since publicly traded stocks are minority shares, if a controlling interest is being valued using publicly listed companies – a control premium should be considered.

In the application of the market approach, several adjustments of the ratios might be necessary to reflect differences in the comparables, degree of marketability, and/or liquidity.

- Timing differences between market transactions and the valuation date.
- Strategic or investment value issues.
- Size, depth of management, diversification of markets, products and services, and relative growth and risk.

102.12.3 Asset-Based Valuation

The asset-based approach to valuation is based on the financial accounting concept that the owners' equity is determined by subtracting the book value of a company's liabilities from the book value of its assets (NACVA 2019). Under the generally accepted accounting principles ("GAAP"), most assets on the balance sheet are recorded at historical cost, whereas most long-term liabilities are recorded at the present value, and there is no recognition of intangible assets. Because of the accounting approach, the "book value" of the owners' equity will normally be underestimated. However, the book value could be adjusted to fair market value, by deriving the fair market value of each asset and liability on the balance sheet of the subject company or entity, and then subtract the total adjusted liabilities from the total adjusted assets to derive the adjusted book value. This approach increases the valuation cost and complexity, but it still ignores intangible assets if they are not listed on the balance sheet. This method does not address the operating earnings of the business, and it would not be inappropriate to value a company or entity with intangible assets.

Under GAAP, held-to-maturity securities may not be reflected at fair value and may need to be adjusted. The first-in first-out ("FIFO") method of inventory usually provides the best indicator of fair value for inventory. The value of plant, property, and equipment is typically not at fair value and will need an appraisal if the amounts are significant. If the amount is not significant, an adjusted depreciation using reasonable useful lives and a non-accelerated depreciation method may result in an acceptable estimate of fair value. Other assets may also need a separate appraisal, depending on their nature. Liabilities are usually at fair value. But loans from related parties may need to be adjusted to reflect their fair value using market interest rates.

102.12.3.1 Liquidating Value

Another asset valuation methodology is the liquidating value for firms in financial distress or firms whose operating prospects are very cloudy. This method is not appropriate for a going concern but may be appropriate if there is doubt whether the company would remain in business.

Accounting for Mergers and Acquisitions

All business combinations need to be accounted for as acquisitions in accordance with the 2009 Financial Accounting Standards Board ("FASB") Accounting Standard Codification ("ASC") 805, Business Combinations are addressed by ASC 820, Fair Value Measurements and Disclosure (Flood 2017). Prior to June 2001, business combinations in the USA were accounted for by using either the purchase method or the pooling of interests' method (Morse, Hemming LLP 2014). On June 30, 2001, the FASB released Statement of Financial Accounting Standards ("SFAS") No. 141, Business Combinations, requiring that all business combinations be accounted for by the purchase method. On December 7, 2007, the FASB issued SFAS 141R, which applies to all transactions in which an acquirer obtains control of

one or more businesses. SFAS 141R is the outcome of cooperation between the FASB and the International Accounting Standard Board (“IASB”), thus contributing to convergence between the USA and international accounting standards (Stevens et al. 2008).

ASC 805 requires to perform a valuation process and a purchase price allocation (“PPA”) to the fair value of all assets acquired, liabilities assumed, and contingent liabilities. The three valuation approaches are the income approach, the market approach, and the cost approach. If the fair value of net assets acquired is less than the total consideration paid for the acquired entity, the difference is recorded as goodwill.

Goodwill impairment is addressed by FASB ASC 350. Goodwill is impaired when its *implied* fair value is less than its carrying amount.

102.13 Post-merger Integration

Post-merger integration is critical for M&A success or failure whereby the acquirer needs to integrate people, process, technology, and information (Hofmeyer 2019). The PMI process requires good communication and efficient governance and management practices, significant experience and attention to cultural profiles and individuals’ behavior (Reimann 2006). Management teams that run the merged process often fail behind in meeting deadlines, performance targets, and in evaluating the new people (Ashkenas and Francis 2000). Too many mergers fail due to ineffective PMI processes. To avoid such blunders, it is critical to select integration managers who can appreciate the cultural issues involved, handle them personally, and help others deal with them constructively. Guiding the post-merger integration process takes a special leader who can handle complex circumstances rapidly and successfully.

Three main post-merger integration models include preservation, absorption, and symbiosis (Haspeslagh and Jemison 1991). Preservation selects the best practice of both companies. This approach has limited integration needs, but it may be insufficient for such activities as changing financial reporting requirements for the parent company and policies. In the absorption approach, the acquirer incorporates, or absorbs, the acquired company into their business by utilizing transitioning people, process, technology, and data, and benefiting from economies of scale and cost cutting. Symbiosis selects the best practices of both the target and acquired companies. Although this approach has the greatest potential to achieve synergies over time, it requires significant change of management and increase the level of risk. This type of integration process can take up to a year or longer to complete and has significant effect on day-to-day operation of the business.

102.14 M&A Risk and Success

Companies often fail to achieve the financial objectives and operational synergies that were projected at the time of deal consummation (Crowe Horwath 2013). Failure occurred when acquirers were overoptimistic in their assumptions; acquirers

overestimated expected synergies; acquirers overbid in the heat of the bidding process; and poor post-acquisition integration. Merger performance could be addressed from different perspectives, including:

- Macro benefits of mergers – for the entire economy/sectors, improved efficiency, and operational performance.
- Micro effects are companies / sectors engaged in mergers benefit from enhanced efficiency or redistribution effect.
- Shareholders' value effect following a deal and over time.

The evidence on M&A performance is mixed, as cited below:

- Success and failure of M&As was defined in terms of *the economic profit model* whether $(\% \text{ econ profit} / \text{invested capital in the acquisition}) > \text{WACC}$; where $\text{econ profit} = \text{noplat} - \text{invested capital} \times \text{WACC}$, or $\text{econ profit} = \text{EBIT} (1 - t) - \text{invested capital} \times \text{WACC}$ (Copeland 2000).
- McKinsey & Company's study of 116 mergers between 1972 and 1982 shows that 61% were failure, 23% were successful, and 16% were not accounted for; a more recent survey by McKinsey of 160 acquisitions in 1995–1996 shows that only 12% of the acquirers managed to accelerate their growth over the next 3 years, and 42% lost ground (McKinsey Quarterly 2001).
- Bain's worldwide research project of more than 1,600 publicly traded companies and more than 18,000 deals from 2000 through 2010 shows that M&A was an essential part of successful strategies for profitable growth. During this period, total shareholders return (TSR) average 4.5% per year, compared with 3.3% for companies that were inactive in M&A.
- Prof Bruner states that some 60–70% of all M&A transactions compensate their investors for their opportunity cost (Bruner 2004).
- A research study of mergers by 276 Fortune 500 companies over a 15-year period shows that more than two-thirds resulted in increased efficiency (measured by cost-per-unit of revenue) and huge savings without significantly reducing employment (Trimbath 2002).

A Deloitte survey of 1000 US executives in late 2016 provides insights into factors contributing to success and impediments to successful mergers, as shown in Table 8.

102.15 Closing Comments

Merger waves were observed since the late nineteenth century that were affected by economic and financial conditions, technology, infrastructure, and regulations. M&As cover a wide spectrum of transaction types that are complex and quite risky, and require the services of various professionals, of which the bulk are

Table 8 Impediments to achieving successful M&As

	% of respondents
Insufficient due diligence process	88
Improper target identification	83
Not valuing the target accurately	81
Changing regulatory and legislative environment	81
Failure to effectively integrate	78

Source: Deloitte, M&A Trends, year-end report 2016, survey of 1000 US executives, Sept. 2016

procured from outside the corporations. Post-merger integration is vital, and its efficient application could assure successful deals.

M&As are significant corporate investment and restructuring activities that are valued in the context of the DCF and market multiples. When the choice is organic investment versus external investment through mergers and acquisitions, the decision is usually in favor mergers.

M&As are subject to intensive regulations related to securities issuing, antitrust, and corporate governance that are applied worldwide. Funding of mergers are through cash, stocks, and a combination of both, with implications on risk sharing between the acquirers and targets.

The Covid-19 crisis might have prompted the evolution of alternative transactions that could reduce risk and required investments.

References

- Abbott, Alden F. 2005. A brief overview of American antitrust law, The University of Oxford Centre for Competition Law and Policy, The Competition Law & Policy Guest Lecture Programme Paper.
- Arzac, Enrique R. 2008. *Valuation for mergers, buyouts and restructuring*. 2nd ed. Wiley.
- Ashkenas, Ronald N., and Suzanne C. Francis. 2000. Integration managers: Special leaders for special times. *Harvard Business Review*.
- Bain & Company. 2014. The renaissance in mergers & acquisitions.
- Berk, Jonathan, and Peter DeMarzo. 2014. *Corporate finance*. 3rd ed, 233–260. Pearson, Ch 8.
- Bertrand, Olivier, and Marie-Ann Betschinger. 2011. *Performance of domestic and cross-border acquisitions: Empirical evidence from Russian acquirers*. European Bank for Reconstruction and Development, Working Paper No. 129.
- Brealey, Richard A., Stewart C. Myers, and Franklin Allan. 2020. *Principles of corporate finance*. 13th ed, 832. McGraw Hill Education.
- Brotherson, W. Todd, Kenneth M. Eades, Robert S. Harris, and Robert C. Higgins. 2014. Company valuation in mergers and acquisitions.
- Bruner, Robert F. 2002. Does M&A pay? A survey of evidence for the decision-maker. *Journal of Applied Finance* 12 (1): 48–68.
- Bruner, Robert F. 2004. *Applied mergers and acquisitions*. University edition. Wiley Finance. Ch. 3, pp XV, 72–75, 325.
- . 2005. *Deals from hell: M&A lessons that rise above the ashes*, 13, 65, 341. Wiley.
- Bruner, Robert F., Robert M. Conroy, Javier Estrada, Mark Kritzman, and Wei Li. 2002. *Introduction to valuation in emerging markets*. Darden Graduate Business School, University of Virginia.

- CFIUS. <https://www.treasury.gov/resource-center/international/Pages/Committee-on-Foreign-Investment-in-US.aspx>; The U.S. Treasury Department, 4, 2014, Issued 2016. The Committee on Foreign Investment in the United States.
- Coeurdacier, Nicolas, Roberto A. De Santis, and Antonin Aviat. 2009. *Cross-border mergers and acquisitions financial and institutional forces*. European Central Bank. Working paper series No 1018.
- Coffee, John C., and Darius Palia. 2014. *The impact of hedge fund activism: Evidence and implications 14*. Working paper No. 266.
- Crowe Horwath. 2013. "Critical pillars for M&A success" survey while corporate deal-makers are optimistic, recent M&A results are disappointing.
- Damodaran, Aswath. 2001. *Corporate finance, theory and practice*. 2nd ed, 840–853. Wiley.
- . 2006. *Damodaran on valuation*, Security analysis for investment and corporate finance. 2nd ed. Wiley.
- . 2012. *The dark side of valuation, valuing young, distressed, and complex businesses*. 2nd ed. FT Press.
- Deloitte. 2020. The state of the deal, M&A trends 2020. <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/mergers-acquisitions/us-mna-trends-2020-report.pdf>
- DePamphilis, Donald M. 2012. *Mergers, acquisitions and other restructuring activities*. 6th ed. Academic. Ch. 3.
- di Giovanni, Julian. 2003. *What drives capital flows? The case of cross-border M&A activity and financial deepening*. Department of Economics, University of California Berkeley.
- Downey, Jim, and Technical Information Service. December 2008. *Mergers and acquisitions*, Topic Gateway Series No. 54.
- Duff & Phelps. 2015. *2015 valuation handbook, guide to cost of capital*. Wiley.
- Eiteman, David K., Arthur I. Stonehill, and Michael H. Moffett. 2016. Chapter 18 Multinational capital budgeting and cross-border acquisitions. In *Multinational business finance*, 14th ed. Pearson.
- Elie, Paul, Bobby Bono, and Barry Misthal. 2016. *Global industrial manufacturing M&A deals insights*. PricewaterhouseCoopers.
- Erel, Isil, Rose C. Liao, and Michael S. Weisbach. 2012. Determinants of cross-border mergers and acquisitions. *Journal of Finance* LXVII (3), 1045–1082.
- Ernst & Young. Technology capital confidence barometer: Drive for innovation and growth sustained technology M&A.
- EY Global Capital Confidence Barometer. 2020. How do we find clarity in the mid of a crisis? https://www.ey.com/en_gl/ccb/advanced-manufacturing-mergers-acquisitions
- Fernandez, Pablo, et al. 2016. *Market risk premium used in 71 countries in 2016: A survey of 6,932 respondents*. Madrid: IESE Business School, University of Navarra. <http://www.valuewalk.com/wp-content/uploads/2016/05/SSRN-id2776636.pdf>
- Fishman, Jay E., Shannon P. Prat, and William J. Morrison. 2007. *Standards of value: Theory and applications*, 21–26. Wiley.
- Flood, Joanne M. 2017. *Wiley GAAP 2017: Interpretation and application of generally accepted accounting principles*. Wiley. ASC 805 Business Combinations; and Ch. 52, ASC 820 Fair Value Measurements ("Wiley GAAP"). [Electronic resource].
- Gaughan, Patrick A. 2011. *Mergers, acquisitions, and corporate restructuring*. 5th ed. Wiley.
- . 2015. *Mergers, acquisitions, and corporate restructuring*. 6th ed. Wiley. Ch. 5.
- Goedhart, Marc, Tim Koller, and David Wessels. 2010. *The five types of successful acquisitions*. McKinsey & Company, Web.
- Goetzmann, William N., and Roger G. Ibbotson. 2006. *The equity risk premium, essays and explorations*. Oxford University Press.
- Hall, Richard. 2014. United States of America takeover guide, Cravath, Swaine & Moore LLP.
- Harroch, Richard, Hari Raman, and Albert Vanderlaan. 10 key questions and answers about SPACs. *Forbes*, Nov 11, 2020., <https://www.forbes.com/sites/allbusiness/2020/11/11/10-key-questions-and-answers-about-spacs/?sh=3b8b47c52f83>

- Hart-Scott-Rodino Annual Report Fiscal Year 2015.
- Haspeslagh, Philippe C., and David B. Jemison. 1991. *Managing acquisitions: Creating value through corporate renewal*. The Free Press.
- Hofmeyer, Stefan. 2019. Chapter 16: Postmerger integration: Lessons from experience and research in successful corporations.
- Holloway, Brian. 2013. Intangible assets in purchase price allocations. *Insights*. www.willamette.com
- Holmström, Bengt R., and Steven Neil Kaplan. 2001. Corporate governance and merger activity in the U.S.: Making sense of the 1980s and 1990s. <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/mergers-acquisitions/us-mergers-acquisitions-trends-survey-the-future-m-a.pdf>
- Humphery-Jenner, et al. 2012. *Cross-border mergers and acquisitions: The role of private equity firms*. Duisenberg School of Finance – Tinbergen Institute Discussion Paper.
- Kengelbach, Jens, et al. 2020. The 2020 M&A report: Alternative deals gain traction, BCG.
- Klausner, Michael. 2003. Investors' choices: Institutional shareholders, private equity, and anti-takeover protection at the IPO stage. *University of Pennsylvania Law Review* 152: 755, 757, 760.
- . 2013. The empirical revolution in law: Fact and fiction in corporate law and governance. *Stanford Law Review* 65: 1325, 1348.
- Kling, Lou R., and Eileen T. Nugent. 2015. *Negotiated acquisitions of companies, subsidiaries and divisions* § 16.06.
- Koerner, Jens. The M&A process revisited – Identifying a suitable phase model, Faculty of Business and Economics, Mendel University in Brno, 613 00 Brno, Czech Republic.
- Koller, Tim, Marc Goedhart, and David Wessels. 2010. *Valuation, measuring and managing the value of companies*. 5th ed. McKinsey & Company, Wiley.
- KPMG. 1999. Unlocking shareholders value: The keys to success, M&A, A Global Research Report.
- Lathan & Watkins. 2020. Acquiring a US-public-company, an overview for the non-US acquirer.
- Layne, Ramey, Brenda Lenahan, and Vinson & Elkins LLP. Special Purpose Acquisition Companies: An introduction. *Harvard Law School Forum on Corporate Governance*, July 6, 2018. <https://www.corpgov.Law.harvard.edu/2018/07/06/special-purpose-Acquisition-companies-an-introduction/>
- Lipton, Martin. 1987. Corporate governance in the age of finance corporatism. *University of Pennsylvania Law Review* 136 (1): 30.
- Lipton, Martin, and Erica H. Steinberger. 2015. Takeovers & freezeouts, § 6.03[2].
- Luehrman, Timothy A. 1997. Using APV: A better tool for valuing operations. *Harvard Business Review*.
- Malmberg, Derek, et al. Private-company CFO considerations for SPAC transactions, September 2020, 1–6, <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/audit/us-private-company-CFO-considerations-for-SPAC-transactions.pdf>
- Malmendier, Ulrike, et al. 2015. Target revaluation after failed takeover attempt, cash v. stock. SSRN.
- Martin, Reimann, et al. 2006. International post-merger integration. In *Mergers & acquisitions management*, ed. B.W. Wirtz, 931–956. Wiesbaden: Gabler. SSRN.
- McKinsey Quarterly. 2014. Why mergers fail? #4.
- Monks, Robert A.G., and Nell Minow. 2011. *Corporate governance*. 5th ed, 182. Wiley.
- Morgan, J. P. 2017. M&A global outlook: Finding opportunities in a dynamic market.
- Morse, Hemming LLP. 2014. Purchase price allocations under ASC 805: A guide to allocating purchase price for business combination. <http://www.hemming.com/wp-content/uploads/2014/08/Purchase-Price-A>
- NACVA. Chapter Six, Commonly used methods of valuation.
- Nelson, Cody, and Towers Watson. 2015. *The changing landscape of golden parachutes in a say-on-pay world*, 1. Willson Towers Watson. <https://www.towerswatson.com/en-US/Insights/Newsletters/Global/executive-pay-matters/2015/Executive-Compensation-Bulletin-The-Changing-Landscape-of-Golden-Parachutes-in-a-Say-on-Pay->
- Notop, Bruce. 2007. Rules to acquire by. *Harvard Business Review*.

- Outline of legal aspects of mergers & acquisitions in the U.S., 15th ed., Sept 2013, Overview of state securities laws.
- Palisade, Monte Carlo Simulation. http://www.palisade.com/risk/monte_carlo_simulation.asp.
- Pereiro, Luis E. 2002. *Valuation of companies in emerging markets: A practical approach*. Wiley.
- Poniachek, Harvey A., ed. 2019. *Mergers and acquisitions: A practitioner's guide to successful deals*. World Scientific.
- Pratt, Shannon P. 2001. *Business valuation discounts and premiums*. Wiley. Ch. 1.
- PWC. 2020. Tomorrow's deal dynamics: How new ways of being are reshaping M&A, Web.
- Rappaport, Alfred, and Sirower. 1999. Stock or cash? Trade-offs for buyers and sellers in mergers and acquisitions. *Harvard Business Review*, 148.
- Renneboog, Luc, et al. 2017. Leveraged buyouts: Harvard Law School Forum on Corporate Governance, an overview of the literature. <https://www.corpgov.Law.Harvard.edu/2017/02/page/06/>
- Reuters. 2008. Hostile takeovers hit record as market swoons. Available at <http://www.reuters.com/article/us-mergers-hostiles-idUSTRE48S2P120080929>
- Rosenbaum, Joshua, and Joshua Pearl. 2013. *Investment banking, valuation, leveraged buyouts, and mergers and acquisitions*. 2nd ed. Wiley.
- Ross, Stephen A., Randolph W. Westerfield, Jeffrey Jafee, and Bradford D. Jordan. 2013a. *Corporate finance*. 11th ed. McGraw-Hill Education. Ch. 4, 9.
- . 2013b. *Corporate finance*. 11th ed. McGraw-Hill Education. p. 243. Ch. 4, 9.
- Ross, Stephen A., Randolph W. Westerfield, Jeffrey F. Jaffe, and Bradford D. Jordan. 2019. *Corporate finance*. 12th ed. McGraw Hill Education.
- Rossi, Stefano, and Paolo Volpin. 2002. *Cross-country determinants of mergers and acquisitions*. London Business School.
- Schill, Michael J., and Elena Loutschina. 2017. Business valuation in mergers and acquisitions. Darden Case No. UVA-F-1699.
- Sharkrepellent.Net. 2008. Rethinking the role of the poison pill?. Available at https://www.sharkrepellent.net/pub/rs_20080916.html
- Shleifer, Andrei, and Robert W. Vishny. 1987. *Management buyouts as a response to market pressures*. University of Chicago Press.
- Solomon, Steven Davidoff. 2013. Men's wearhouse dusts off the Pac-Man defense. *The New York Times*. http://dealbook.nytimes.com/2013/11/26/mens-wearhouse-dusts-off-the-pac-man-defense/?_r=0
- SSRN: <https://ssrn.com/abstract=261112> or <https://doi.org/10.2139/ssrn.261112>.
- State of Delaware, Title 8, Corporations, Chapter 1. General corporation law, Subchapter IX. Merger, consolidation or conversion. <http://delcode.delaware.gov/title8/c001/sc09/index.shtml>
- Stevens, Snell & Wilmer L.L.P. 2008. Get ready for FAS 141(R), new accounting rules for business combinations.
- Sudarsanam. 2003. *Creating value from M&A*. FT Prentice Hall.
- Surane, Jennifer, and Crystal Tse. The Reddit crowd also loves SPACs. *Bloomberg BusinessWeek*, February 1, 2021.
- Thompson, Robert B. 1999. Preemption and federalism in corporate governance: Protecting shareholder rights to vote, sell and sue. *Law and Contemporary Problems* 62: 215.
- Thomson, Reuters. 2008. *Mergers & acquisitions review*, 5.
- Trimbath, Susanne. 2002. *Mergers and efficiency, changes across time*. Springer.
- Tuttle, Douglas, and Deloitte. 2020. Merging alliances: Ignoring at your own risk. [IMAA-Institute.org](https://www.imaa-institute.org)
- U.S. Department of Justice and the Federal Trade Commission. 2010. Horizontal merger guidelines. <https://www.justice.gov/atr/horizontal-merger-guidelines-08192010>
- Wachtell, Lipton, Rosen & Katz. 2016. Takeover law and practices. <http://www.wlrk.com/files/2016/TakeoverLawandPracticeGuide.pdf>
- Weston, J. Fred, Mark L. Mitchell, and J. Harold Mulherin. 2004. *Takeover, restructuring, and corporate governance*. 4th ed. Pearson Prentice Hall.



Accrual Accounting and Risk: Abnormal Sales Growth, Accruals Quality, and Returns

103

Liu Min Shirley

Abstract

This study proposes a new business risk proxy, unexpected sales growth (AG), which is measured as the difference between a firm's sales growth rate in year t and its benchmark, the weighted average of sales growth rate over the past 3 years. I find that the AG is statistically, economically, significantly associated with other risk measures of risk (e.g., beta, volatility of cash flows, operating cycle, and negative earnings occurrences). I find that the more the sales growth of a firm in the current year deviates from its benchmark – the higher business risk the firm – the lower the accrual quality, and the higher future abnormal returns the firm, which support Penman's (2016) theoretical conclusion that accrual accounting is for risk.

Keywords

Abnormal Sales Growth · Operating Risk · Accruals Quality · Future Returns

JEL Classification

G10 · G11 · M40 · M41

103.1 Introduction

This study is motivated by the expected important relationship between accruals quality and uncertainty embedded in the sales growth, a type of business risk, and the lack of empirical evidence on this association. Previous studies test and find that the volatility of sales inversely influences the accruals quality. However, if one could perfectly predict the sales growth, one could perfectly estimate the changes in the

L. Min Shirley (✉)

Department of Accounting, University of Illinois, Chicago, IL, USA

e-mail: mliu5@fau.edu

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_106

2501

accounts related to the sales. In other words, the estimable sale volatility would not increase the estimation errors in the accruals for the accounts related to sales given other conditions fixed, consequently, would not deteriorate the accruals quality. Therefore, I argue that the unexpected (hereafter abnormal) sale growth, which may cause the estimation errors in the accruals of the accounts related to sales, thereby, lowering accruals quality. Since the abnormal sales growth could cause the uncertainty in future earnings (or cash flows) which could infer the firm's value, even when the firm's profit margin is fixed, a risk that needs to be compensated if it can't be diversified away, therefore, I would expect that the more deteriorated sales growth is associated with a higher rate of returns.

Abnormal sales growth (hereafter AG) is measured as the difference between a firm's sales growth rate in year t and its weighted average sales growth rate over the preceding 3 years, which serves as a benchmark or expected value for sales growth in the current year. Accounting textbooks explain that changes in a firm's total accruals are associated with changes in sales relevant accounts such as accounts payable and accounts receivable. If sales grow at a constant rate over time, the sales relevant accounts are also stable. If there are unexpected changes in the sales growth rate, the amount of accruals in sales relevant accounts would fluctuate and be difficult to predict, thereby, resulting in unexpected accruals. Dechow and Dichev (2002) and McNichols (2002) measure the accruals quality as the standard deviation of residuals obtained from regressing working capital accruals on cash from operations in the current period, prior period, and future period (while controlling for change in revenues and plant, property, and equipment). Therefore, I hypothesize that the residuals from the accruals regression should be related to AG based on the relationship between unexpected sales growth rate and unexpected accruals.

The abnormal sales growth rate proposed in this study conceptually differs from the volatility of sales in Francis et al.'s (2005) modified Jones (1991) model. The volatility of sales in their paper reflects fluctuation in the level of sales. AG reflects fluctuation in changes in sales growth rate. Francis et al. (2005) find that the total accruals quality of a firm depends more on a firm's business risk than on managerial discretion. I conjecture that AG captures an additional dimension of business risk that goes beyond just changes in sales. The empirical results are consistent with this expectation. Adding AG as an explanatory variable to the Francis et al. (2005) regression of total accruals quality on economic fundamental variables (size, volatility of cash flows, volatility of sales, operating cycle, and occurrence of negative earnings over past 10 years) increases the R-square of the model by nearly ten percentage points.

This paper also investigates whether the impact of AG on total accruals quality depends on the sign of abnormal sales growth. I hypothesize that for firms recently experiencing negative AG, an increase in AG would increase total accruals quality. This effect manifests because these firms' economic fundamentals are approaching a steady state (less uncertainty of future earnings) and tend to perform better. On the other hand, for firms recently experiencing positive AG, their accruals quality should decrease with further increases in AG because changes in economic fundamentals may cause more volatility in the firms' business operations – higher level uncertainty

for future earnings. The empirical results presented later in this chapter are consistent with these expectations.

Penman (2010, page 670) and Ohlson (2014) reason that growth risk is one of fundamental risks that is inherent in the business of the firm. Sales can drive the growth in net operating assets, which influences residual earnings, the driver of value of the firm. Moreover, the “growth risk is driven by the risk of sales not growing as expected.” Therefore, I expect that abnormal sales growth shall influence the value of the firm. Although this theory is very convincing, there is a lack of empirical evidence on the relationship between unexpected sales growth rate and valuation. To examine the relationship between abnormal sales growth rate and abnormal returns, using monthly return data, I form ten portfolios sorted on AG ranks in the preceding year and test the significance of the Jensen’s-alpha for each portfolio using the Fama and French four-factor model. Results of the four-factor model show that portfolios in the two-lowest and two-highest tiles of decile sorted by AG ranks have higher abnormal returns and larger beta (i.e., higher systematic risk). The portfolio with the lowest AG rank in the preceding year has a, on average, 18.48% abnormal returns in the current year, suggesting that the returns are compensated to riskier assets. By investigating the relationship between AG and future abnormal returns, this study adds empirical evidence to Penman’s valuation theory about the uncertainty ingrained in the abnormal sales growth and firm values.

Desai et al. (2004) suggest trading strategies, which combine the level of total accruals and sales growth. This study is different from Desai et al. (2004) in a few ways. First, this study uses the changes in sales growth rate as opposed to the nominal level of sales growth. Additionally, this study incorporates the measure of accruals quality suggested by Francis et al. (2005) and not the total accruals level. Second, I estimate abnormal returns measured as Jensen’s-alpha, the intercept of the Fama-French four factor regression model using the monthly return series. Desai et al. (2004) use size-adjusted returns as the abnormal return variable.

To examine whether and how the firms’ characteristics change with the changes in AG of the firms, I form ten portfolios, which are sorted basing upon AG ranks. For portfolios where AG is greater than zero, the firms’ profitability, sales, and assets nearly monotonically decrease with an increase in AG. Firms with slightly positive AG have a large size, moderate profitability (i.e., return on equity and return on assets) and market to book ratio (PB), and slightly slower sales growth rate (8.59%) compared with that of a full sample (13.84%). On the other hand, for portfolios where AG is less than or equal to zero, I find that firms’ profitability, sales, and assets nearly monotonically increase with an increase in AG.¹ These results suggest that when a firm approaches its steady state, it has a smaller magnitude of sales growth rate, large assets, large sales, large net income, a negative external financing ratio, (which could be indicated by the reduced the firm’s common equity), the highest total accruals quality, (which is measured as in Francis et al. 2005), and lower market

¹A firm with slightly negative (or close to zero) AG has relatively stable sales growth, almost no fluctuation in sales growth.

risk (proxied by beta). These findings are consistent with Ohlson's (2008) analysis of accruals and growth and offer a possible explanatory factor to the accrual anomaly suggested by Ohlson and Bilinski (2015).

The empirical evidence collectively suggests that the measure of abnormal sales growth rate proposed by this study can be a proxy for uncertainty ingrained in sales growth, a type of operating risk, driving the accruals quality and the abnormal returns, which compensate investors for bearing higher risk (i.e., operating /growth risk and accruals quality/information risk). This study extends the current literature by providing first-hand empirical evidence on the effects of abnormal sales growth on accruals quality and future returns and on how the characteristics of firms change with the change in abnormal sales growth rate. The results in this chapter are robust to sensitivity tests such as alternative measures of abnormal returns and various econometrics methods.

This chapter proceeds as follows. Section 103.2 reviews relevant literature and develops hypotheses. Section 103.3 outlines the methods used to test the hypotheses. Section 103.4 describes the samples and presents descriptive statistics for the variables used in my analyses. Section 103.5 reports the empirical results. Section 103.6 presents the results of robust tests and additional analyses. Section 103.7 concludes.

103.2 Background and Hypotheses Development

103.2.1 Abnormal Sales Growth and Accruals Quality

The Generally Accepted Accounting Principles (GAAPs) require that firms prepare their financial statements basing upon the accruals accounting system because, in a short interval of time, the accrual accounting reflects a firm's real business operating and the true value of the firm compared with the cash-based accounting system. A cash-based accounting system recognizes revenues when the firm receives cash, and records expenses when cash is paid out. Accruals accounting requires firms to book revenues when a transaction finishes and the amount of the cash related to the transaction and the time of receiving cash can be reasonably estimated.

On one hand, the accrual accounting system is more appropriate than the cash-based accounting system. It is quite common that a vendor may not be able to receive the cash immediately after shipping the goods to its customers. If this scenario happens to the firm at its fiscal year end, shall this firm record the revenues for the goods shipped to its customers even if it has not received the cash yet? The cash-based accounting system does not recognize the revenues until the firm receives the cash from its customers. But the accruals accounting system recognizes the revenues if the amount of cash and the time of receiving cash related to these transactions can be reasonably estimated. Therefore, accruals accounting reflects the actual time when transactions really happen. Thus, accruals accounting is more accurate in terms of reflecting a firm's real business operating compared to cash-based

accounting. Dechow (1994) discusses the circumstances, under which accruals are a better measurement of a firm's performance than cash.

On the other hand, human judgment is involved in the process of estimating the amount of receivables and timing of receiving cash – estimation of accruals. The managers of firms may manage earnings by manipulating accruals such as booking revenues when it should not book and write off a large amount of bad debt when it is inappropriate to do so. Therefore, the good and bad sides of accruals accounting could play a role in the financial reports. Thus, it is important to know whether a firm's financial reports have a high or low accruals quality.

DeAngelo (1986) uses total accruals from a prior period ($t-k$) as a measure of the “normal” total accruals. Then, DeAngelo (1986) defines the “abnormal” total accruals (ΔTA_t) as the difference between current total accruals and normal total accruals, which can be separated into discretionary and nondiscretionary accrual components.

$$\Delta TA_t = (TA_t - TA_{t-k}) = (DA_t - DA_{t-k}) - (NA_t - NA_{t-k}) \quad (1)$$

DeAngelo assumed that the average change in nondiscretionary accruals ($NA_t - NA_{t-k}$) is approximately zero, then a change in total accruals ($TA_t - TA_{t-k}$) primarily reflects a change in discretionary accruals ($DA_t - DA_{t-k}$). This assumption is too restricting. Jones (1991) relaxes this assumption.

To estimate the expected total accruals, Jones (1991) suggests a regression model that total accruals is a function of a change in revenues and gross property, plant, and equipment in year t for firm j . The unexplained residuals (ϵ) from the following model, Eq. (2), are the proxies for the level of discretionary component of accruals. Total accruals include changes in working capital accounts (i.e., accounts receivable, inventory, and accounts payable) that depend to some extent on changes in revenues. Therefore, Jones (1991) controls for change in revenues to reflect the economic environment of the firm. Gross property, plant, and equipment are also included to control for the portion of total accruals related to the nondiscretionary depreciation expenses.

$$TCA_{j,t} = \beta_{0,j} + \beta_{1,j}Rev_{j,t} + \beta_{2,j}PPE_{j,t} + \epsilon_{j,t} \quad (2)$$

Jones (1991) does not consider one role of accruals that is to shift or adjust the recognition of cash flows over time so that the adjusted earnings better measure firm performance (Dechow and Dichev 2002) (hereafter DD). DD propose the following firm-level time-series regression model, Eq. (3), mapping the working capital accruals in the current period to cash flows in the prior, current, and future periods. The residuals ($\epsilon_{j,t}$) of Eq. (3) are the portion of total working capital accruals unrelated to realized cash flows. The standard deviation of these residuals is a firm-level measure of accruals quality.

$$\Delta WC_{j,t} = \beta_{0,j} + \beta_{1,j}CFO_{j,t-1} + \beta_{2,j}CFO_{j,t} + \beta_{3,j}CFO_{j,t+1} + \epsilon_{j,t} \quad (3)$$

McNichols (2002) combines the DD model and the Jones (1991) model and suggests that the total working capital accruals are the function of cash flows in

prior, current, and future periods, and changes in revenues and in gross property, plant, and equipment in year t for firm j . The new combined model increases the adjusted R-square of the DD regression model, which indicates an improvement in the prediction of total accruals.

Francis et al. (2005) further refine the combined model suggested by McNichols (2002). Francis et al. (2005) use the industry level estimated total accruals as the benchmark of total accruals for a firm j at year t . The standard deviation of the residuals of the regression, Eq. (4), of total accruals on the cash flows in prior, current, and future periods, and changes in revenues and in gross property, plant, and equipment in year t for firm j over years $t-4$ through t is the proxy for the total accruals quality (hereafter AQ) for the firm j at year t . The larger the AQ is, the lower the accruals quality is.

$$TCA_{j,t} = \phi_{0,j} + \phi_{1,j}CFO_{j,t-1} + \phi_{2,j}CFO_{j,t} + \phi_{3,j}CFO_{j,t+1} + \phi_{4,j}\Delta Rev_{j,t} + \phi_{5,j}PPE_{j,t} + v_{j,t} \quad (4)$$

DD identified five innate factors affecting accruals quality: firm size (Size), standard deviation of cash flows from operations (STDCFO), standard deviation of sales revenues (STDSale), length of operating cycle (OperCycle), and incidence of negative earnings realizations (NegEarn). Extending DD, Francis et al. (2005) argue that total accruals quality (AQ) can be broken down to innate accruals quality reflecting a firm's business environment and discretionary accrual quality affected by the management. The fitted value from the annual regression of Eq. (5) is the innate accruals quality and residuals of the Eq. (5) are discretionary accruals that could be a proxy for earnings management.

$$AQ_{j,t} = \lambda_0 + \lambda_1 Size_{j,t} + \lambda_2 STDCFO_{j,t} + \lambda_3 STDSales_{j,t} + \lambda_4 OperCycle_{j,t} + \lambda_5 NegEarn_{j,t} + \mu_{j,t} \quad (5)$$

Following Francis et al.'s (2005) argument that the innate factors of accruals quality reflect a firm's business environment, I argue that a change in sales growth rate should affect the firm's accruals quality because the sales growth rate can reflect the level of operating risk, a type of business risk, of the firm (Ohlson 2014; Penman 2010). The change in sales growth rate can represent the change in a firm's level of operating (business) risk, thereby, reflecting the change in the business environment of the firm.² Thus, a change in sales growth rate in period t should be associated with the accruals quality currently or subsequently.

² Accounting textbooks show that the amount of accruals is associated with the amount of sales. The more sales a firm has in a given year t , the larger amount of receivable and payable associated with the sales the firm has. One would expect that the fluctuation in sales would cause fluctuation in accruals amount, therefore, impact on accruals quality as well. Not all firms with a change in sales growth have comparable accounting system that can capture changes in accruals related to change in sales growth immediately. Then, these firms may have unexplained accruals called discretionary accruals (Jones 1991), or accruals quality (Francis et al. 2005).

Furthermore, theoretically, for the firms with steady growth rate ($AG=0$), the accrual quality decreases with either an increase or a decrease in the growth rate because any deviation from the steady state may create unexpected accruals associated with abnormal change in the growth rate, which lower the firms' accrual quality given that I define the lower accruals quality as the larger amount unexplainable accruals. Using abnormal sales growth rate as the proxy for the change in sales growth rate and for the change in the level of operating risk, I expect that the unexpected sales growth rate (AG) is related to accruals quality (AQ). My first hypothesis is stated in alternative form below.

H1: The further the sales growth rate of a firm deviates from its expected value, the lower the accruals quality of the firm.

A change in sales growth can be upward ($AG>0$) or downward ($AG<0$). For the purpose of my analyses, I refer to firms with positive AG as high growth firms, firms with negative AG as low growth firms, and firms with a zero AG as steady growth firms. My analyses focus on the high and low growth firms because there is no firm with a zero AG . For the high growth firms ($AG>0$), if sales growth increases, the larger the value of the abnormal sales growth, the more volatile the business operation of the firm is, thereby, the firm is more likely to have more volatile accruals and lower accruals quality (i.e., larger value of the standard deviation of the residuals of the regression Eq. (5) above). Therefore, I expect that, for firms with positive unexpected sales growth, the accrual quality decreases with the increase in the sales growth rate. Thus, for the high growth ($AG>0$) firms, H1 can be stated in the alternative form as following:

H1A: Accruals quality of high growth firms is negatively associated with abnormal sales growth rate.

For the low growth firms ($AG < 0$), if sales growth rate increases, the larger the value of the abnormal sales growth, the less volatile the business operating of the firm is, thereby, the firm is more likely to have less volatile accruals and higher accruals quality. Therefore, I expect that, for firms with negative abnormal sales growth, the accrual quality increases with the increase in the sales growth rate. Thus, for the low growth firms, H1 can be stated in the alternative form below:

H1B: Accruals quality of low growth firms is positively associated with abnormal sales growth rate.

103.2.2 Abnormal Sales Growth and Abnormal Returns

The previous literature suggested the market overprices the high growth firms (e.g., Graham and Dodd 1934; Lakonishok et al. 1994; Doukas et al. 2002; Desai et al. 2004) and investors can long low growth firms (value firms with low M/B, P/E, P/C ratios) and short high growth firms (glamour firms with high M/B, P/E, P/C ratios) to generate hedge profits. Desai et al. (2004) suggested a trading strategy combining the level of total accruals and sales growth rate.

Penman (2010, page 670) says that growth risk is priced by capital market and "... growth risk is driven by the risk of sales not growing as expected." However, there is a lack of empirical research on the relationship between abnormal sales

growth and valuation. Basing on the above discussion that a change in sales growth rate ($AG \neq 0$) represents the change in the firm's operating risk, I expect that abnormal sales growth rate (AG) is associated with a firm's future stock performance. Therefore, my second hypothesis is stated in the alternative form as following.

H2: The further the sales growth rate of a firm deviates from its expected value, the higher future returns of the firm will have.

The analyses on the relationship between abnormal sales growth rate (AG) and future returns for the high and low growth firms are similar to those for the relationship between AG and AQ. Therefore, H2 can be restated below in alternative forms separately for high and low growth firms.

H2A: The future returns of high growth firms positively associated with abnormal sales growth rate.

H2B: The future returns of low growth firms negatively associated with abnormal sales growth rate.

The next section discusses the empirical method used to test the above hypotheses.

103.3 Methods Used to Test Hypotheses

103.3.1 Model Used to Test H1

To capture the effect of unexpected change in sales growth on the accruals quality in the current period, I add abnormal sales growth rate (proxied for the change in business risk or environment) to the regression Eq. (4).

$$AQ_{j,t} = \lambda_0 + \lambda_1 AG_{j,t} + \lambda_2 AGLag1_{j,t} + \lambda_3 Assets_{j,t} + \lambda_4 STDCFO_{j,t} + \lambda_5 STDSales_{j,t} + \lambda_6 OperCycle_{j,t} + \lambda_7 NegEarn_{j,t} + \mu_{j,t} \quad (6)$$

where, $AQ_{j,t}$ is accruals quality. Following Francis et al. (2005), I estimate the regression Eq. (4) cross-sectionally by industry-year. I group firms into the 48 Fama and French (1997) industries. The $AQ_{j,t}$ is estimated as the standard deviation of the residuals of regression Eq. (4) over years t-4 through t for firm j at year t. The larger (smaller) value of AQ is the poorer (better) of the accruals quality. Please refer to the detailed calculation of AQ in the appendix.

$AG_{j,t}$ is the variable of abnormal sales growth for firm j in year t, measured as the difference between sales growth rate for firm j in year t and the benchmark of firm j's sales growth rate, which is measured as the weighted average of sales growth from years t-3 to t-1. $SaleGrowth_{j,t}$ is computed as $\left(\frac{Sales_{j,t}}{Sales_{j,t-1}} - 1 \right)$.

$$AG_{j,t} = SaleGrowth_{j,t} - SalesGBechmark_{j,t}$$

$$SalesGBechmark_t = 0.3 \times SalesG_{t-3} + 0.3 \times SalesG_{t-2} + 0.4 \times SalesG_{t-1}$$

In Sect. 103.2, the H1A predicts that, for high growth firms ($AG > 0$), the firms' abnormal sales growth rate is negatively associated with the firm's accruals quality. Therefore, I expect, for the high growth firms ($AG > 0$) in year t , the coefficient on the variable AG in the regression Eq. (6) to be positive ($\lambda_1 > 0$). The H1B predicts that, for the low growth firms ($AG < 0$), firms' abnormal sales growth rate is positively associated with the firm's accruals quality. Thus, I expect, for the low growth firms ($AG < 0$), the coefficient on the variable AG in the regression Eq. (6) to be negative ($\lambda_1 < 0$).

$AGLag1_{j,t}$ is one period of lag value of abnormal sales growth for a firm j in year t . A change in sales growth in the prior year may have an impact on accruals quality in the current year. The time series impact of AG on accruals quality is unclear and has not been documented in previous literature either. I do not predict the sign of λ_2 .

$Assets_{j,t}$ captures the size effect, measured as the log value of total assets. $STDCFO_{j,t}$ is cash flow variability, measured by the standard deviation of firm j 's rolling 10-year cash flows from operation, scaled by total assets. $STDSale_{j,t}$ is sales variability, measured as the standard deviation of the firm j 's rolling 10-year sales revenues, scaled by total assets. $OperCycle_{j,t}$ is operating cycle, measured by log of the sum of the firm j 's days account receivable and days inventory. $NegEarn_{j,t}$ is the occurrence of negative earnings realization, measured by the firm j 's proportion of losses over the prior 10 years. Following DD and Francis et al. (2005), one may expect that firms with smaller size, larger cash flow volatility, longer operating cycles, and a higher incidence of losses, to have poorer accruals quality. This indicates that $AQ_{j,t}$ is negatively associated with $Assets$ and positively associated with $STDCFO_{j,t}$, $STDSale_{j,t}$, $OperCycle_{j,t}$, and $NegEarn_{j,t}$ in Eq. (6).

103.3.2 Model Used to Test H2

To test my second hypothesis that a firm's abnormal sales growth in current period (t) is associated with its stock performance in future period ($t+1$), the Jensen-alpha approach (or the calendar-time portfolio approach)³ is used to assess the risk-adjusted abnormal returns of portfolios based on the decile of ranked abnormal sales growth in the preceding year.

The four factor model (Fama and French 1993; Carhart 1997) is adopted to compute the abnormal returns of each of the ten portfolios using monthly return data series.

³The Jensen-alpha approach is an alternative to the buy-and-hold abnormal returns calculating the calendar-time portfolio returns for firms in each decile of abnormal sales growth in preceding year, and calibrating whether portfolio returns are abnormal in a multifactor regression (Kothari and Warner 2007, 24-27). I use four-factor model in this study.

$$R_{p,t} - R_{f,t} = \alpha_{p,t} + \beta_{p,t} (R_{m,t} - R_{f,t}) + s_{p,t} SMB_t + h_{p,t} HML_t + u_{p,t} UMD_t + \varepsilon_{p,t} \quad (7)$$

Where,

$(R_{p,t} - R_{f,t})$ = the portfolio risk-adjusted return in month t ;

$(R_{m,t} - R_{f,t})$ = the risk-adjusted market return at time t ;

SMB_t = the size factor at time t , calculated as the difference between the returns on a portfolio of small stocks and a portfolio of big stocks;

HML_t = the book-to-market factor, calculated as the difference between the returns on a portfolio of high book-to-market equity (BE/ME) stocks and a portfolio of low BE/ME stocks.

UMD_t = the stock momentum at time t , representing the difference between two high prior return portfolios and two low prior return portfolios.⁴

In the returns test, the estimated intercept coefficient, $\alpha_{p,t}$, from regression Eq. (7) for each value-weighted portfolio is interpreted as the abnormal returns in excess of what could have been earned by passive investment in these four factors. Each portfolio is rebalanced every fiscal year during the sample period (January 1974 through December 2011) basing on the ranks of AG of sample firms in the preceding year.⁵

Basing on the construct of AG, the proxy for risk, I would expect that the Jensen-alpha is greater than zero ($\alpha_{p,t} > 0$) for the portfolios with the lowest and the largest rank of AG in the preceding year.

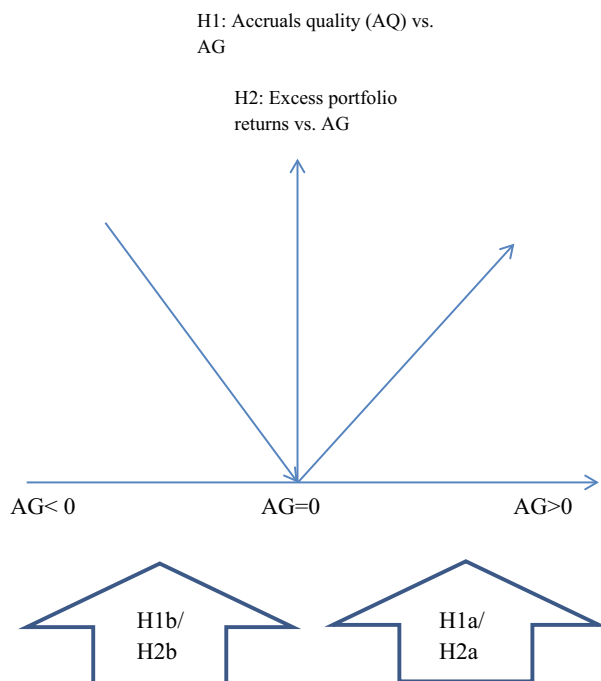
103.4 Sample and Descriptive Statistics

103.4.1 Sample Selection and Data

Table 1 describes the sample selection procedure. I begin from all firms with available annual financial data in Compustat to compute the abnormal sales growth and total accruals for the period of years from 1964 through 2011. Following Francis et al. (2005), the calculation of a firm- and year-specific accruals quality (AQ) needs the previous 9 years of data ($t-1, \dots, t-9$), and the AQ is based on five annual residuals. The sample also excludes firms in financial industry because firms in financial industry have different financial structures from other industry's firms and are under different regulations. Then, the accruals sample is restricted to firms with at least 7 year data and with 178,008 firm-year observations for 15,736 distinct firms. Because the sample selection procedure requires firms with at least 7 year financial data in Compustat and

⁴Thanks to Ken French for providing these factors on his website: <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french>.

⁵In year t , a firm enters into a portfolio based on decile of ranked AG in the next month after its financial reports for prior year released if it had an AG falling into the range of that given decile of ranked AG at the end of year $t-1$, and leaves the portfolio 12 months after entering into it if its AG rank changes in the next year.

Fig. 1 Illustration of hypotheses

at least 12 month stock returns data in CRSP, it may bias my sample to surviving firms, which tend to be larger and more successful than the population. However, I expect that this restriction would decrease the variation in AQ, making it more difficult to detect the effect of environmental factors on AQ, which makes, if there is the sample selection bias, my results should be more reliable. The accruals quality results are available for a 38-year period beginning from year 1973 ending at year 2010.⁶

Because testing my H2 needs monthly return data from CRSP, to keep sample firms containing annual data sample and monthly return data series, I restrict my final sample of annual data to firms with available monthly return data and necessary annual financial data. These procedures lead to the final annual data sample with 103,965 firm-year observations and 10,480 distinct firms.

Because the relationship between AG and AQ are opposite for a firm with a positive change in sales growth and a firm with a negative change in sales growth, I further divide my sample into two subsamples. If a firm has a positive (negative or no) change in sales growth in year t , then, the firm belongs to the high (low) growth subsample. There are 43,683 (60,282) firm-year observations with 8,752 (10,038)

⁶The based year in my sample is year 1964; therefore, the 1st year to have a firm-year observation of AQ is 1973. Because the AQ calculation needs 1-year lead cash flows data, the AQ in year 2010 needs cash flows data from year 2011.

Table 1 Sample selection

Descriptions	Firm-year observations	Distinct firms
Firm-years in Compustat fundamental annual database from 1964 to 2012 (including all observations with available data to compute total accruals and abnormal growth)	341,143	31,695
Less firm-years with		
Missing Compustat data to compute accruals quality	178,008	15,736
And excluding financial institutions		
<u>Accruals Sample</u>	<u>163,135</u>	<u>15,959</u>
Less firm-year without return data in CRSP	59,170	5,479
<u>Final Sample</u>	<u>103,965</u>	<u>10,480</u>

distinct firms in the high (low) growth subsample over the 38-year period (1973–2010).

The next step is to obtain monthly return series from CRSP. For the final annual sample, there are 1,160,905 firm-year-month observations of available monthly returns beginning in January 1974 ending in December 2011. Each value-weighted (equal-weighted) portfolio basing on the ranks of abnormal sales growth in preceding year has 465 time series observations.

To avoid the effects of outliers, I winsorize the data at two tails (1% and 99% of the variable distribution).

103.4.2 Sample Summary Statistics

Table 2 reports the summary statistics for the pooled final sample. The number of firms each year ranges from about 1,700 to 3,800 firms per year,⁷ which is close to Francis et al. (2005) sample firm year distribution.⁸ Mean and median values of AQ are 0.0447 and 0.0323, respectively; 80% of the values are in the range of 0.0044–0.0985. These results are close to those in Francis et al. (2005) (having mean and median values of AQ 0.0442 and 0.0313, respectively, and 80% of the values in the range of 0.0107–0.0943). The slight difference in AQ statistics between my sample and theirs may be caused by the over-time variation in AQ (Francis et al. 2005, 306).

The mean and median values of abnormal sales growth (AG) are -0.1032 and -0.0336 , respectively; 80% of the values are in the range of -0.4746 – 0.2725 .

⁷The earliest year available in my sample is 1973 with 275 firms. However, it is not representative in my sample. Since 1973, there are about 1,700–3,800 firms per year. The sample reaches its largest number of distinct firms per year in 2002 (3,862), then declines over year toward the end of the sample period (2,621 distinct firms in year 2010).

⁸The number of firms each year in Francis et al. (2005) sample ranges from about 1,500 per year in the early 1970s to about 3,500 firms per year toward year 2001.

Table 2 Sample descriptive statistics ^a

Variable	Mean	Median	Std Dev	25%	75%	10%	90%
SALE	1590.66	211.03	4297.84	51.21	928.81	15.63	3565.41
AT	1825.52	194.06	5150.90	48.24	954.74	17.09	3940.10
MCAP	2335.80	157.53	12112.43	33.90	825.54	10.81	3443.54
IB	107.56	5.42	840.10	-0.13	39.83	-12.61	186.70
ROE	0.0813	0.0949	13.8242	0.0031	0.1564	-0.2846	0.2335
SaleGrowth	0.1384	0.0877	0.3745	-0.0165	0.2138	-0.1540	0.4317
AG	-0.1032	-0.0336	0.6408	-0.1902	0.0886	-0.4746	0.2725
AQ	0.0447	0.0323	0.0437	0.0164	0.0587	0.0044	0.0985
InnateAQ	0.0571	0.0438	0.0465	0.0253	0.0745	0.0144	0.1183
DisAQ	-0.0008	0.0000	0.0117	-0.0034	0.0031	-0.0106	0.0087
Size	5.4316	5.2682	2.0759	3.8761	6.8614	2.8384	8.2790
STDCFO	0.1383	0.0808	0.2130	0.0491	0.1400	0.0306	0.2580
STDSale	0.4091	0.2809	0.4250	0.1661	0.4866	0.0944	0.8498
OperCycle	3.8544	3.9922	0.8067	3.6336	4.2889	2.9456	4.5933
NegEarn	0.2314	0.1000	0.2874	0.0000	0.4000	0.0000	0.7000
Leverage	0.2736	0.2319	0.3187	0.0684	0.3890	0.0000	0.5528

^aThe donation of each variable for year t is suppressed in this table.

The sample contains 103,965 firm-year observations for 10,480 distinct firms over the 38-year period (1973–2010).

The summary statistics for AG indicate that the sample firms' sales growth has declined on average over the sample period (1973–2010), but firms in 75th (90th) percentile have their sales growth increased, on average, by 8.86% (27.25%) over the sample period. Untabulated summary statistics show that there is a large variation in AG overtime. The year 2002 (and 2009) has the lowest (next to lowest) sales growth, $AG = -0.30825(-0.276)$ for year 2002 (2009) among the sample period, indicating that the sales growth for all firms declined, on average, by 30.8% (27.6%) in year 2002 (2009). The summary statistics on abnormal sales growth is consistent with the macroeconomic condition over years in the USA. The year 2002 is the year after internet bubble and the economy was slow in that year; the year 2009 is the year post financial crisis of year 2008 (and late 2007). The mean value of AG for year 2010 is 0.018266, indicating that the sales growth of the sample firms in the year 2010 increased by 1.8266% on average, consistent with the economy recovery reported in the press.

Moreover, Table 2 shows the summary information on selected financial variables. The sample firms are large (median market value of equity is about 157.53 million dollars and median assets are about 194.06 million dollars), profitable (median income before extraordinary items are about 5.42 million dollars, median return on equity is 9.49%), and growing (median sales growth is 8.77%).

Furthermore, Table 2 also reports the summary statistics for the innate factors, affecting accruals quality, identified in Dechow and Dichev (2002). The mean and median values of log value of the total assets (Size) are 5.4316 and 5.2682, respectively. The mean and median values of cash flow variability over past

10 years (STDCFO) are 0.1383 and 0.0808, respectively. The mean and median values of sales variability over the past 10 years (STDSale) are 0.4091 and 0.2809, respectively. The values of mean and median of the log value of operating cycle (OperCycle) are 3.8544 and 3.9922, correspondingly. The mean and median values of incidence of negative earnings (NegEarn) over the past 10 years are 0.2314 and 0.1, respectively. The mean and median values of incidence of negative earnings (NegEarn) over the past 10 years are 0.2314 and 0.1, respectively. The mean and median values of ratio of total interest-bearing debt to total assets (Leverage) are 0.2736 and 0.2319, respectively. Overall, compared with Francis et al.'s (2005) sample, the sample firms in this study are more volatile in cash flows and sales over the past 10 years and have more frequent occurrence of negative earnings. These results may be due to inclusion of the 2007/2008 financial crisis period in this sample.

103.5 Empirical Results

103.5.1 Results of Testing H1

103.5.1.1 Correlation Results

High-Growth Subsample

Panel A of Table 3 reports the Pearson (above the diagonal) and Spearman (below the diagonal) correlations for the variables used in a high growth subsample (firms with a positive change in sales growth) analyses.

The Pearson correlation between AG and its one period lag is -0.1092 , statistically significant at less than 0.0001 level, indicating a firm with a positive change in sales growth may have a negative change in sales growth in the prior year. However, the Spearman correlation between AG and its one period lag is 0.0036 but not statistically significant. The Pearson (Spearman) correlation between AG and AQ is 0.2750 (0.2239) and statistically significant at less than 0.0001. Recall the larger AQ, the poorer accruals quality. This result indicates that the accruals quality of a firm with a positive change in sales growth ($AG > 0$) decreases with the increase of abnormal sales growth (AG), consistent with my H1a in Sect. 103.2..

Panel A of Table 3 also reports that the Pearson (Spearman) correlation between the abnormal sales growth (AG) and the log value of the total assets (Size) is -0.16 (-0.1749) and statistically significant at less than 0.0001, indicating that a firm with a large amount of sales growth tends to be smaller in size, which is consistent with conventional information that small firms are usually high growth firms. The Pearson (Spearman) correlation between abnormal sales growth (AG) and volatile of cash flows (STDCFO) and volatile of sales (STDSale) are 0.2745 and 0.1741 (0.2926 and 0.2076), respectively, both of which are statistically significant at less than 0.0001, indicating that the volatile of cash flows and of sales increases with the increased change in sales growth for a firm with a positive change in sales growth in year t . The Pearson (Spearman) correlation between abnormal sales growth

(AG) and external financing (XFIN) is 0.2858 (0.2005), statistically significant at less than 0.0001, indicating that the likelihood of external financing increases with the increased change in sales growth for a firm with a positive change in sales growth in year t . This result is consistent with rational expectation that a firm more likely needs external financing when it experiences high sales growth.

Low-Growth Subsample

Panel B of Table 3 reports the Pearson (above the diagonal) and Spearman (below the diagonal) correlations for the variables used in low growth subsample (firms with a negative change in sales growth) analyses.

The Pearson (Spearman) correlation between AG and its one period lag is 0.4245 (0.2106), statistically significant at less than 0.0001 level, indicating a firm with a negative change in sales growth may have a negative change in sales growth in prior year. The Pearson (Spearman) correlation between AG and AQ is -0.1617 (-0.1848) and statistically significant at less than 0.0001, indicating that the accruals quality for low growth firms, where AG is less or equal to zero ($AG \leq 0$), increases with the increase of abnormal sales growth (AG), consistent with my H1b in Sect. 103.2.

Results in panel B of Table 3 also show that the Pearson (Spearman) correlation between abnormal sales growth (AG) and log value of the total assets (Size) is 0.1714 (0.2085) and statistically significant at less than 0.0001, indicating that a large firm's sales growth tends to approach its steady stage, which is consistent with conventional information that large firms usually are mature firms and have stable sales growth. The Pearson (Spearman) correlation between abnormal sales growth (AG) and volatile of cash flows (STDCFO) and of sales (STDSale) are -0.3384 and -0.2021 (-0.3750 and -0.2562), respectively, both of which are statistically significant at less than 0.0001, indicating that the volatile of cash flows and of sales decreases with the increasing sales growth for a firm with a negative change in sales growth in year t . The Pearson (Spearman) correlation between abnormal sales growth (AG) and external financing (XFIN) is -0.2048 (-0.1193), statistically significant at less than 0.0001, indicating that the likelihood of external financing decreases with increasing sales growth for a firm with negative change in sales growth in year t . This result is still consistent with rational expectation that a firm less likely needs more external financing when its sales growth approaches a steady state.

These results for the low growth subsample are generally opposite to those for the high growth subsample. While supporting my H1 that the AG is associated with the accruals quality and firm's business risk or the business environment volatile, those results are consistent with my H1A that a firm's business risk (or the business environment volatile) increases with the increasing sales growth for the firm with a positive change in sales growth (more volatile sales growth), and with my H1B that a firm's business risk (or the business environment volatile) decreases with the increasing sales growth for the firm with a negative change in sales growth (lower volatile sales growth).

Table 3 Pearson (above the diagonal) and Spearman (below the diagonal) correlations for variables in analyses. (I suppress the subscripts j, t of each variable in the table.)

Panel A: correlation table for the high-growth subsample ^{a,c}												
Variable	AG	AGLag1	AQ	Size	STDCFO	STDSale	OperCycle	NegEarn	ROE	ROA	PB	XFIN
AG		-0.1092	0.2750	-0.1600	0.2745	0.1741	0.0517	0.3619	0.0013	-0.1935	0.0055	0.2858
AGLag1	0.0036		-0.0561	0.0394	-0.0957	-0.0462	-0.0259	-0.1179	0.0014	0.1082	0.0043	-0.0388
AQ	0.2239	-0.0255		-0.3114	0.3995	0.3286	0.0961	0.4188	0.0392	-0.1475	0.0494	0.1773
Size	-0.1749	0.0210	-0.3038		-0.2775	-0.2580	-0.0663	-0.3575	-0.0103	0.1490	-0.0165	-0.1248
STDCFO	0.2926	-0.0284	0.5597	-0.4753		0.5022	0.0441	0.4185	0.0852	-0.1257	0.0898	0.1879
STDSale	0.2076	-0.0076	0.4062	-0.3721	0.6144		-0.1154	0.2481	0.0242	-0.0309	0.0275	0.0959
OperCycle	0.0667	-0.0249	0.1320	-0.1081	0.1419	-0.0643		0.0811	-0.0079	-0.0802	-0.0035	0.0578
NegEarn	0.2925	-0.0411	0.4335	-0.3595	0.5063	0.3234	0.1115		-0.0240	-0.4234	-0.0018	0.2950
ROE	-0.0935	0.1354	-0.1249	0.1856	-0.1378	-0.0597	-0.1278	-0.4038		0.2385	0.8665	-0.0419
ROA	-0.1160	0.1330	-0.1157	0.0906	-0.0836	-0.0287	-0.1051	-0.4554	0.8276		0.2051	-0.3470
PB	0.0950	0.0371	0.1554	0.0805	0.1653	0.0889	0.0505	0.0539	0.2723	0.2753		-0.0326
XFIN	0.2005	0.0207	0.0913	-0.0936	0.1500	0.1037	0.0716	0.1809	-0.2060	-0.2654	0.0577	
Panel B: correlation table for the low-growth subsample ^{b,c}												
Variable	AG	AGLag1	AQ	Size	STDCFO	STDSale	OperCycle	NegEarn	ROE	ROA	PB	XFIN
AG		0.4245	-0.1617	0.1714	-0.3384	-0.2021	-0.1007	-0.4145	0.0091	0.2184	-0.0032	-0.2048
AGLag1	0.2106		-0.1186	0.0872	-0.1956	-0.1188	-0.0216	-0.2355	0.0085	0.1156	-0.0004	-0.0809
AQ	-0.1848	-0.1173		-0.2924	0.3887	0.3169	0.1238	0.3498	0.0081	-0.1534	0.0229	0.1273

Size	0.2085	0.0684	-0.2590		-0.2858	-0.2371	-0.1110	-0.3704	0.0067	0.1887	-0.0047	-0.1405
STDCFO	-0.3750	-0.1488	0.5283	-0.4724		0.4788	0.0746	0.4509	0.0514	-0.1529	0.0936	0.1979
STDSale	-0.2562	-0.1138	0.3872	-0.3481	0.6163		-0.0835	0.2241	0.0143	-0.0490	0.0323	0.0535
OperCycle	-0.2133	-0.0387	0.1507	-0.1585	0.1840	-0.0284		0.1340	-0.0038	-0.1165	0.0004	0.0713
NegEarn	-0.3647	-0.1547	0.3460	-0.3732	0.5209	0.2935	0.1680		-0.0256	-0.3838	-0.0035	0.3064
ROE	0.2888	0.1503	-0.1582	0.2685	-0.2272	-0.1291	-0.1898	-0.5123		0.0958	0.7373	-0.0241
ROA	0.3239	0.1657	-0.1736	0.2262	-0.2129	-0.1127	-0.1825	-0.5982	0.8401		0.0678	-0.2814
PB	-0.0244	0.0078	0.0840	0.0776	0.1305	0.0485	0.0486	0.0212	0.1844	0.2479		-0.0133
XFIN	-0.1193	-0.0156	0.0354	-0.0971	0.1259	0.0510	0.0775	0.1910	-0.1989	-0.2466	0.0597	

^aResults in panel A of Table 3 are based on a high-growth subsample of 43,683 firm-year observations and 8,752 distinct firms over the 38-year period (1973–2010).

^bResults in panel B of Table 3 are based on a low-growth subsample of 60,282 firm-year observations and 10,038 distinct firms over the 38-year period (1973–2010).

^cThe bolded correlation coefficients are statistically significant at less than 5% level.

103.5.1.2 Regression Results

This subsection reports two types of regression results of testing H1A and H1B. I first use the F test in the ordinary least square (OLS) regressions to validate the model, Eq. (6), examining whether AG is a good explanatory variable for AQ. Then, I use the Fama and MacBeth (1973) annual regression method to further test H1.

Ordinary Least Square Regressions

Panels A and B of Table 4 shows F tests for the validation of seven regression models with AG as an independent variable, using high growth and low growth subsample, respectively. Each model adds one more independent variable (I.V.) to the prior regression model. The F test statistics for the associated change in R-squares, when an I.V. is added to the previous model, indicates whether the newly added I.V. is a suitable explanatory variable for the dependent variable. A significant increase in the R-squares associated with newly added I.V. can be interpreted as that the newly added I.V. is a good explanatory variable for the dependent variable.⁹ Therefore, whether increment in R-square associated with the newly added AG, the I.V., is statistically significant (shown by F test statistics) can infer whether abnormal sales growth (AG) is a good explanatory variable to total accruals quality (AQ).

The model 1 in Panel A of Table 4 shows that 9.74% of the variation of accruals quality (AQ) can be explained by only one independent variable AG, and the F statistics for this model is 3572.92, indicating that the model 1 is statistically valid. The estimated coefficients of AG are range from 0.0160 (in the model 7) to 0.0372 (in the model 1) and are all statistically significant at 0.01 levels, consistent with my H1A that the accrual quality decreases with an increase in the sale growth rate.

The next to the last column in Panel A of Table 4 shows that, in model 3, the variable Size contributes most to the change in R-square (11.01% increment to model 2). The variable of one period lag of AG also can explain a portion of variation in AG, but the sign of the coefficient of one period lag of AG flipped from negative (−0.0019) to positive (0.0017), when adding more control variables identified by Dechow and Dichev (2002). The results in Table 4 suggest that the negative sign of the coefficient of AGLag1 is likely due to the omitted correlated independent variables (e.g., Size, STDCFO). The sign of the coefficient of AGLag1 was corrected once including the omitted independent variables.

Using the low-growth subsample, the analyses for the results in Panel B of Table 4 are similar to those, using high-growth subsample, reported in Panel A of Table 4. The estimated coefficients of AG range from −0.0076 (in the model 7) to −0.0241 (in the model 1), indicating that accrual quality increases with an increase in sales growth rate. The result supports my H1B.

⁹Here, the current model is the full model, while the prior model without the newly added independent variable is the nested model. When an independent variable is added to the nested model, the R-square changes. If the R-square of the current model is significant, shown by F-test statistics, larger than that of the prior model, I can infer the newly added independent variable is a good explanatory variable for the dependent variable (Greene 2008, 81–92).

Table 4 The F test results of changes in R-squared of the OLS regression of accruals quality on abnormal sales growth

Panel A: High-growth subsample ^a

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		
	Predict Sign	Coef Est	t-stat	Coef Est	t-stat	Coef Est	t-stat	Coef Est	t-stat	Coef Est	t-stat	Coef Est	t-stat	Coef Est	F-stat
Intercept	?	0.0374	153.16	0.0374	153.18	0.0740	126.23	0.0617	101.56	0.0566	88.78	0.0375	31.82	0.0275	24
AG	+	0.0372	59.8	0.0370	59.39	0.0313	53.09	0.0241	41.47	0.0236	40.93	0.0232	40.39	0.0160	28.26
AGLag1	?		-0.0019	-3.44	-0.0013	-2.53	0.0007	1.39	0.0007	1.43	0.0009	1.84	0.0017	3.64	0.0974
Size	-				-0.0063	-67.86	-0.0051	-55.63	-0.0047	-52.08	-0.0046	-50.49	-0.0033	-36.48	0.2079
STD CFO	+							0.0576	53.83	0.0447	37.73	0.0426	36.06	0.0254	21.48
STD Sale	+									0.0129	24.63	0.0147	27.68	0.0143	28.1
OperCycle	+											0.0047	19.16	0.0042	17.67
NegEarm	+													0.0442	52.99
															0.0553

Panel B: Low-growth subsample ^b

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		
	Predict Sign	Coef Est	t-stat	Coef Est	t-stat	Coef Est	t-stat	Coef Est	t-stat	Coef Est	t-stat	Coef Est	t-stat	Coef Est	F-stat
Intercept	?	0.0404	188.83	0.0404	188.67	0.0792	146.75	0.0662	119.17	0.0603	104.03	0.0379	35.95	0.0279	27.03
AG	-	-0.0241	-63.87	-0.0240	-58.48	-0.0198	-51.13	-0.0136	-35.49	-0.0132	-34.78	-0.0123	-32.56	-0.0076	-20.44
AGLag1	?			-0.0002	-0.47	0.0001	0.47	0.0016	5.21	0.0017	5.8	0.0015	5.21	0.0019	6.59
Size	-					-0.0067	-77.42	-0.0054	-63.74	-0.0051	-60.11	-0.0048	-57.26	-0.0034	-40.63
STD CFO	+							0.0548	63.34	0.0422	44.77	0.0403	42.88	0.0242	25.48
STD Sale	+									0.0143	31.71	0.0160	35.48	0.0165	37.85
OperCycle	+											0.0054	25.24	0.0046	22.43
NegEarm	+													0.0429	57.05
															0.0500

^a. Results in panel A of table 3 are based on a high-growth subsample having 43,683 firm-year observations and 8,752 distinct firms over a 38-year period (1973-2010).

^b. Results in panel B of table 3 are based on a low-growth subsample having 60,282 firm-year observations and 10,038 distinct firms over a 38-year period (1973-2010).

The bolded coefficients are statistically significant at less than 5% level. Please refer variable definitions to Table 106.2.

The goal of using the F test in the OLS regression in this subsection is to validate whether AG is a good explanatory variable for AQ. Therefore, I omit the details of other regression analyses here. Next subsection reports the Fama and MacBeth (1973) method regression results and details the implication of regression results of examining the relationships between abnormal sales growth (AG) and accruals quality (AQ).

Fama-MacBeth (1973) Regressions

High-Growth Subsample Panel A of Table 5 reports the results of estimating Eq. (6) for the high-growth subsample ($AG > 0$). The based model is the model 1 as Francis et al. (2005). Coefficients of all independent variables in the based model have the same signs as those reported in Francis et al. (2005). The based model has an adjusted R-square of 0.307, slightly lower than Francis et al.'s (2005), which may be due to different sample periods in two studies.¹⁰

The model 2 in Panel A of Table 5 adds the variable of abnormal sales growth (AG) to the based model (Dechow and Dichev 2002; Francis et al. 2005). The coefficient on AG is 0.011, statistically significant at less than 0.01 level with a 9.25 of the Fama-Macbeth t-statistics, and 20.37 (9.87) of Z_1 (Z_2) test statistics (Fama and MacBeth 1973),¹¹ indicating that the AQ increases by 0.011 unit with one-unit increase in the AG for a high-growth firm in year t holding other factors constant. The result that the variable AG is still statistically significant at less than 0.01 level and have a relatively large magnitude (0.011) in the regression with the presence of standard deviation of sales suggests that the variable AG captures an aspect of innate (fundamental) determinants of AQ, which is different from the volatile of sales (STDSale). The untabulated results of testing multicollinearity show that there is no multicollinearity issue in this regression model.¹² The adjusted R-square of the model 2 increased to 0.319 from 0.307 of the based model, indicating that the variable AG still can explain a good amount of variation in AQ after controlling for traditional innate accruals factors (i.e., size, volatiles of

¹⁰This study uses firms in Compustat in later years than Francis et al. (2005). The average accounting quality of Compustat sample firms in recent years are generally lower than that of sample firms in earlier years because Compustat has expanded its database and included more small firms, which usually have a lower accounting quality than large firms.

¹¹Here, FM.T-stat is the Fama Macbeth t-statistics computed as $FM.T - stat = \frac{\bar{t}\sqrt{N-1}}{stdev(t)}$. Z_1 and Z_2 statistics test whether the time-series mean t-statistics from yearly regressions is statistically different from zero, computed as $Z_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{t_i}{\sqrt{k_i(k_i-2)}}$, $Z_2 = \frac{\bar{t}}{stdev(t)\sqrt{N-1}}$, where t is t-statistic and k is the degrees of freedom for year i , and N is the number of years.

¹²All VIFs are less than 10.

cash flows and sales, operating cycle, and occurrence of negative income).¹³ The untabulated results also show that the root of the mean of squared error (MSE) of the model 2 is less than that of the based model. Since the model with the minimum MSE is a better choice than other models, results of the model 2 suggest that the AG is one of the omitted variables in accruals quality regression models.

The model 3 in Panel A of Table 5 examines the effect of one period lag value of AG on the accruals quality controlling for the traditional accruals innate factors and AG. The coefficient on the one period lag of AG is 0.001 and not statistically significant at 5% levels in the FM t-test and Z_2 test statistics of time series t-stat. The adjusted R-square of the model 3 increased to 0.355 from that of the model 2 (0.319). The VIF and Tolerance tests show there is no multicollinearity issue in the model 3. However, the untabulated OLS regression results basing on robust t-statistics show that the coefficient of one period lag of AG is statistically significant at 0.000 level. The prior period change in sales growth may affect the next period total accruals (such as account receivables and account payables, and even the cash flows). Therefore, I decide to leave the AGLag1 in the model 3.

Coefficients on all traditional innate factors of accruals quality in all of three regression models have the signs same as predicted and documented in previous studies (e.g., DD; Francis et al. 2004 and Francis et al. 2005): size (Size) is negatively related to AQ; volatile of cash flows and sales in past 10 years, operating cycle, and occurrence of negative earnings (i.e., STDCFO, STDSale, OperCycle, and NegEarn, respectively) are positively associated with AQ.

Notes for panels A and B:

- (1) Z_1 and Z_2 statistics test whether the time-series mean t-statistics from yearly regressions is statistically different from zero.

$$FM.T - stat = \frac{\bar{t}\sqrt{N-1}}{stdev(t)}, \quad Z_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{t_i}{\sqrt{k_i(k_i-2)}}, \quad Z_2 = \frac{\bar{t}}{stdev(t)\sqrt{N-1}}, \quad \text{where } t \text{ is}$$

t-statistic and k is the degree of freedom for year i, and N is the number of years (N=38 in this table).

- (2) The bolded coefficients are statistically significant at less than 5% level.

Low-Growth Subsample Panel B of Table 5 reports the results of estimating Eq. (6) for the low-growth subsample ($AG < 0$). Similar to the previous analyses for the high-growth subsample, I still focus on the analyses on the effect of AG on AQ. The model 2 in Panel B of Table 5 adds the variable of abnormal sales growth (AG) to the based model. The coefficient on AG is -0.003 , statistically significant at 0.05 level with a -2.35 of Fama and Macbeth t-statistics, and $-6.56(-2.05)$ of Z_1

¹³ The magnitude of incremental R-square by adding an independent variable is related to the order of adding that independent variable. And also the regression method used in this subsection is annual regression different from OLS regression used in previous analyses.

Table 5 Results of regression of accruals quality on abnormal sales growth effect using Fama and MacBeth method

Panel A: High-growth subsample ^a													
Independent		Model 1: Based model				Model 2: AG vs. AQ				Model 3: AG and AGLag1 vs. AQ			
Variable	Expected Sign	Coef.	FM. T-stat	Z1	Z2	Coef.	FM. T-stat	Z1	Z2	Coef.	FM. T-stat	Z1	Z2
Intercept	?	0.028	9.7	25.62	10.62	0.028	9.87	25.37	10.74	0.028	10.2	25.92	10.36
AG	+					0.011	9.25	20.37	9.87	0.013	9.87	24.42	11.84
AGLag1	?									0.001	1.27	3.11	1.54
Size	-	-0.003	-12.94	-34.48	-13.67	-0.003	-12.9	-34.87	-13.65	-0.003	-15.1	-37.33	-13.99
STDCFO	+	0.044	7.79	30.1	9.32	0.041	7.43	27.53	8.67	0.049	6.76	29.23	8.68
STDSale	+	0.013	10.16	25.51	10.74	0.012	10.14	24.6	10.84	0.013	11.62	25.86	11.22
OperCycle	+	0.004	9.07	16.22	9.69	0.003	8.88	15.84	9.51	0.004	10.23	17.07	10.62
NegEarn	+	0.038	17.58	46.9	17.35	0.034	16.99	41.13	17.24	0.037	20.45	43.07	16.27
Yearly Obs.		930				930				895			
Adj. R2		0.307				0.319				0.355			

Panel B: Low-growth subsample ^b

Independent Variable	Expected Sign	Model 1: Based model				Model 2: AG vs. AQ				Model 3: AG and AGLag1 vs. AQ			
		Coef.	FM. T-stat	Z1	Z2	Coef.	FM. T-stat	Z1	Z2	Coef.	FM. T-stat	Z1	Z2
Intercept	?	0.028	9.58	27.28	10.71	0.028	9.52	27.46	10.51	0.03	10.52	29.41	10.9
AG	–					–0.003	–2.35	–6.56	–2.05	–0.009	–7.43	–20.67	–8.11
AGLag1	?									0.002	2.75	5.91	2.81
Size	–	–0.003	–12.98	–37.84	–15.67	–0.003	–12.86	–38.03	–15.24	–0.004	–14.68	–41.41	–15.67
STDCFO	+	0.035	6.11	28.21	10.49	0.034	5.98	26.7	10.52	0.041	7.25	30.47	13.17
STDSale	+	0.015	13.03	33.73	15.55	0.015	14.23	34.33	17.3	0.017	15	35.92	15.81
OperCycle	+	0.004	11.45	21.94	12.27	0.004	10.77	21.26	11.56	0.004	10.56	21.12	11.18
NegEarn	+	0.035	20.79	46.12	24.15	0.034	18.74	43.56	23.63	0.038	20.13	46.78	20.04
Yearly Obs.		1198				1198				1152			
Adj. R2		0.288				0.294				0.353			

^aResults in panel A of Table 3 are based on a high-growth subsample of 43,683 firm-year observations and of 8,752 distinct firms over the 38-year period (1973–2010).

^bResults in panel B of Table 3 are based on a low-growth subsample of 60,282 firm-year observations and of 10,038 distinct firms over the 38-year period (1973–2010).

(Z_2) test statistics (Fama and MacBeth 1973),¹⁴ indicating that the AQ decreases 0.003 units with one-unit increase in the AG for a low-growth firm in year t , holding other factors constant. The results indicate that, for low-growth firms, the variable AG captures a dimension of innate determinants of AQ, which is different from volatile of sales (STDSale). The untabulated results of testing multicollinearity show that there is no multicollinearity issue in this regression model and model 3. The adjusted R-square of model 2 increased to 0.294 from 0.288 of the based model, indicating that the variable AG still can explain some variation in AQ after controlling for traditional innate accruals factors.

Results of the model 3 in Panel B of Table 5 suggest that the variable of one period lag of AG is a good explanatory variable for AQ. The coefficient on $AGLag1$ is 0.002, statistically significant at 0.01 level with a 2.75 of Fama and Macbeth t -statistics, and 5.91 (2.81) of Z_1 (Z_2) test statistics (Fama and MacBeth 1973). The adjusted R-square of the regression model 3 increased to 0.353 from 0.294, that of model 2.

Similar to the results in Panel A of Table 5, the coefficients of all traditional innate factors of accruals quality in all of three models have the same signs as those documented in previous literature.

103.5.2 Results on Ranked AG Portfolio Returns (H2)

Panel A of Table 6 reports results of regression Eq. (7), four-factor model (Fama and French 1993; Carhart 1997), where the dependent variable, $(R_{p,t} - R_{f,t})$, is the monthly-return difference between value-weighted portfolio returns and risk-free rate in that month. Each portfolio is constructed based on the rank of AG in the preceding year and rebalanced annually. T-test statistics are reported below the corresponding estimated coefficients for a 456-month time series regression of each portfolio. Therefore, the intercept estimate coefficient, α , is interpreted as the abnormal returns in excess of what could have been earned by the passive investment in these four factors.

The largest monthly abnormal returns are 0.0154 (0.0082) for the lowest (highest) sales growth decile sample firms, statistically significant at 0.01 level. These results are not surprising because the glamour and value stock phenomenon have been well documented in finance and accounting literature. However, the interesting questions are why investing these lowest sales growth firms gains the highest excess returns, and whether the excess returns were earned because of underpricing the low-growth firms or compensating for bearing the higher risk of these low-growth firms.

¹⁴ Here, FM.T-stat is the Fama Macbeth t -statistics computed as $FM.T - stat = \frac{\bar{t}\sqrt{N-1}}{stdev(\bar{t})}$. Z_1 and Z_2 statistics test whether the time-series mean t -statistics from yearly regressions is statistically different from zero, computed as $Z_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{t_i}{\sqrt{k_i(k_i-2)}}$, $Z_2 = \frac{\bar{t}}{stdev(\bar{t})\sqrt{N-1}}$ where, t is t -statistic and k is the degrees of freedom for year i , and N is the number of years.

Recall H1 predicts that a large change in sales growth rate is associated with higher information risk (proxied by AQ) no matter which direction the change is. And the results analyzed in Sect. 103.5.1 support H1 (H1A and H1B). Therefore, the excess returns for lower/higher sales growth portfolios may be related to firms' business risk and information risk that can't be diversified away. Then, beta, the proxy for systematic risk, may help us to disentangle the low/high sales growth excess return puzzle.

The results of estimating beta in column (2) in Table 6 show that the bottom decile of the AG ranked portfolios has the largest beta (1.1631) compared with other deciles of AG ranked portfolios, and statistically significant at 0.001 level, and the 2nd lowest sales growth portfolio has the 2nd largest beta (1.1096) statistically significant at 0.001 level. The 3rd (4th) largest estimated beta, 0.9964 (0.987) is from the highest (2nd highest) sales growth portfolio. The monthly excess returns for the highest (2nd highest) sales growth portfolio is 0.0082 (0.0081). The smallest (second smallest) beta is 0.9019 (0.9280) for the smallest (second smallest) change in sales growth portfolio with the AG rank of 6 (5). The results on portfolio excess returns and beta, at least partially, support the argument that the higher excess return earned by the portfolios with the largest change in (upward or downward) sales growth is required by investors for bearing higher risk. These results are consistent with H2A and H2B that AG is related to the future returns in the predicted directions, and further support that the new risk proxy suggested by this study, AG, is associated with excess returns and the systematic risk beta.

Size effect on excess returns has been well documented. Results in this study support the size effect on excess returns too. However, only portfolios with a big swing in sales growth rate can harvest abnormal returns by investing small firms. The coefficients on SMB for the lowest, second lowest sales growth portfolios in the four-factor regressions are 0.5113 and 0.3118, statistically significant at 0.01 levels. Similarly, the coefficient on SMB for the highest sales growth portfolio is 0.1884, statistically significant at 0.01 level.

The coefficients on HML (book-to-market factor) are generally negative, which may be due to portfolios constructed by the ranks of changes in sales growth that capture the growth factor as well. The coefficients on momentum factor (UMD) do not have a special pattern as those of beta coefficients.

The panel B of Table 6 reports results of abnormal returns for equal-weighted portfolios, which are qualitatively similar to those of value-weighted portfolios. However, the excess return pattern in equal-weighted portfolios is weaker than that in value-weighted portfolios. The pattern for the systematic beta is the same as that in the value-weighted portfolio. The size effect on the equal-weighted portfolios is more pronounced than that on the value-weighted portfolios. This result is not surprising because I do not control for the market capital of sample firms in the equal-weighted portfolios. Results reported in Panel B of Table 6 are consistent with my H2.

In the above portfolio tests, after controlling for growth factor (HML) in the four-factor regressions, the statistically significant Jensen-alpha further suggests that abnormal sales growth captures the different dimension of business risk from traditional growth proxies. And this result is also economically meaningful that

Portfolios by AG rank		(1)	(2)	(3)	(4)	(5)	Adj. R ²
		$a_{p,t}$	$R_{m,t} - R_{ft}$	SMB _t	HML _t	UMD _t	
1	Coef. Est	0.0036	1.1335	1.3675	-0.0514	-0.3499	0.8354
	t-stat	2.07	29.39	24.72	-0.87	-9.19	
2	Coef. Est	0.0026	1.0939	1.0641	0.1414	-0.3279	0.768
	t-stat	1.44	26.44	17.89	2.23	-8.05	
3	Coef. Est	0.0034	1.0248	0.8916	0.1973	-0.2518	0.9082
	t-stat	3.5	47.63	28.89	6.01	-11.85	
4	Coef. Est	0.0036	0.9805	0.8353	0.2767	-0.2181	0.9244
	t-stat	4.45	54.02	32.08	9.98	-12.17	
5	Coef. Est	0.0031	0.9943	0.6852	0.3218	-0.1391	0.9442
	t-stat	4.69	68.31	32.81	14.47	-9.68	
6	Coef. Est	0.0032	0.9643	0.6686	0.2986	-0.1271	0.9355
	t-stat	4.69	63.17	30.53	12.81	-8.43	
7	Coef. Est	0.0041	0.9899	0.6311	0.2603	-0.1102	0.7591
	t-stat	2.81	30.26	13.42	5.2	-3.42	

(continued)

Table 6 (continued)

Panel A: Value-weighted portfolios based on the AG ranks in preceding year						
Portfolios by AG rank		(1)	(2)	(3)	(4)	(5)
		$a_{p,t}$	$R_{m,t} - R_{f,t}$	SMB _t	HML _t	UMD _t
8	Coef. Est	0.0036	0.6942	0.2392	-0.0924	0.9228
	t-stat	4.75	28.29	9.16	-5.47	
9	Coef. Est	0.0035	0.7896	0.1428	-0.0968	0.9118
	t-stat	3.89	27.7	4.7	-4.93	
10	Coef. Est	0.0012	1.0721	-0.0434	-0.1240	0.8814
	t-stat	0.96	39.78	-1.05	-4.66	
$R_{p,t} - R_{f,t}$						

= the portfolio risk-adjusted return in month t;
 $(R_{m,t} - R_{f,t})$
= the risk-adjusted market return at time t;
 SMB_t
= the size factor at time t, calculated as the difference between the returns on a portfolio of small stocks and a portfolio of big stocks;
 HML_t
= the book-to-market factor, calculated as the difference between the returns on a portfolio of high book-to-market equity (BE/ME) stocks and a portfolio of low BE/ME stocks.
 UMD_t
= the stock momentum at time t, representing the difference between two high prior return portfolios and two low prior return portfolios
Adj. R² denotes the adjusted R-square for the regression.
The bolded coefficients are statistically significant at less than 5% level.
The *Italian* number below a coefficient is the t-statistics for that coefficient.

one can earn 18.48% abnormal annual returns by holding firms with the lowest AG rank in the preceding year.¹⁵

In a sum, one can see that correlation and regression results support H1 and H2: the larger change in sales growth rate (either upward $AG > 0$, or downward $AG < 0$) is associated with higher business risk, poorer accruals quality, and larger future returns. This result partially overlaps with but is different from the findings in Liu (2021), which documents that abnormal sales growth rate is negatively associated with the 1-year-ahead size-decile adjusted return but positively associated with unexpected current-period earnings. The results of Liu (2021) and this study indicate that the capital market is rational in terms of not being over optimistic about the sales growth by counting the growth glamor effect in the estimated value of the firm, therefore, requiring higher return to compensate the risk embedded in the sales growth rate in both downward and upward direction.

103.6 Additional Tests

103.6.1 Robust Tests

103.6.1.1 Firm-Specific, Industry, and Year Effects

To address the firm-specific effect, I also run the AQ regressions on AG using the fixed effect (by firms) model (i.e., Eq. (8) below).

$$AQ_{j,t} = \lambda_0 + \lambda_1 AG_{j,t} + \lambda_2 AGLag^1_{j,t} + \lambda_3 Assets_{j,t} + \lambda_4 STDCFO_{j,t} + \lambda_5 STDSales_{j,t} + \lambda_6 OperCycle_{j,t} + \lambda_7 NegEarn_{j,t} + \sum_{t=1973}^{2010} b_{t-1973} Y_t + \sum_{i=2}^{48} i_{i-1} Ind_{j,t} + \mu_{j,t} \quad (8)$$

where $\sum_{t=1973}^{2010} b_{t-1973} Y_t$ are the year dummy variables for years 1973–2010; and $\sum_{i=2}^{48} i_{i-1} Ind_{j,t}$ are the industry dummy variables, where industries are classified as in Fama and French (1997).

The results of the regression Eq. (8) using firm fixed effect are qualitatively same as those reported in Tables 4 and 5, with larger R-squares. That might be due to that adding year and industry dummy variables could explain more variations in AQ.¹⁶

¹⁵Results of the value-weighted portfolio tests show that Jensen-alpha is 1.54% per month, then, will be 18.48% (1.54% x 12) annually.

¹⁶Results of regression Eq. (8) using firm-fixed effect regression method are not tabulated here, and available upon request.

103.6.1.2 Unexpected Sales Growth (AG) and Volatility of Sales.

To investigate the extent to which the abnormal sales growth (AG) contributes to the accruals quality model (i.e., Eq. (8)) compared to volatility of sales documented in Francis et al. (2005), I compute the standardized coefficients on the abnormal sales growth (AG) and volatility of sales (*STDSales*) in the regression Eq. (8). The untabulated results show that the coefficients on AG and *STDSales* are quantitatively same as those reported in terms of magnitude of absolute values of coefficients and statistical significance levels.¹⁷

103.6.1.3 Alternative Econometric Methods

To ensure that the reported results are robust to different econometric estimate methods, I repeat the analysis for Tables 5, clustering the standard errors by the firm to account for possible correlation of regression residuals (Petersen 2009). The estimated coefficients are qualitatively same, with corresponding t-statistics larger than those obtained by Fama and MacBeth (1973) regression procedure.

103.6.1.4 Firm-Level Abnormal Return

I also rerun the regression Eq. (7) at the firm level by the ranks of AG in the preceding year. The results of firm-level regressions are qualitatively similar to those of value-weighted and equal-weighted portfolio-level regressions, but with smaller adjusted R-squares.

103.6.2 Additional Tests

Section 103.5 analyzes the relationship between unexpected sales growth (AG) and accruals quality (AQ) and future returns. Results in Table 6 provide some explanation for the abnormal returns of the lowest and highest sales growth portfolios, so to speak, AG represents the change in operating risk, and therefore, it is associated with AQ, abnormal returns, and beta.

This subsection further disentangles the relationships between certain firms' characteristics and AG. Table 7 reports mean value of each variable of each portfolio, which is sorted based on the decile ranks of AG in the current year.

Table 7 shows that mean values of sales, total assets, market capital, income before extraordinary items increase in the increasing in the AG ranks when moving from the rank 1 to the rank 6, then decrease with the increasing in the ranks of AG, when moving from the rank 6 to the rank 10. The mean values of sales growth increase monotonically along with the increase in AG ranks. The portfolios with the AG rank of 5, 6, and 7 have firms with almost no change in sales growth, suggesting that these firms are about in the steady stage in terms of sales growth.

¹⁷The standardized coefficients on the AG and *STDSales* are 0.135 and 0.148, respectively, and statistically significant at 0.001 levels for the high-growth firms (AG > 0). Results are available upon request.

Table 7 Summary statistics for the firm characteristics based on the rank of abnormal sales growth^a

AGRank	Sales	AT	Meap	IB	SaleGrowth	AG	AQ	Ret_f	PB	OperCycle	NegEam	ROE	ROA	Leverage	XFIN	DEquity	Ddebt
1	483.44	713.18	738.77	-15.12	-0.0607	-1.4514	0.0636	0.0844	3.9789	4.1485	0.4959	-0.4496	-0.1811	0.2697	0.0707	0.0574	0.0109
2	1136.60	1430.07	1679.87	54.74	-0.0555	-0.3841	0.0529	0.0248	8.7072	4.0296	0.2902	0.3124	-0.0549	0.2475	0.0130	0.0115	0.0026
3	1536.74	1788.69	2312.21	100.30	0.0031	-0.2115	0.0455	0.0444	4.1203	3.9063	0.2206	0.0304	-0.0037	0.2532	-0.0012	-0.0026	0.0035
4	1705.85	1907.61	2324.66	112.89	0.0373	-0.1226	0.0401	0.0644	2.5743	3.7960	0.1791	-0.0371	0.0175	0.2554	-0.0067	-0.0100	0.0046
5	2003.03	2199.71	2718.72	130.89	0.0619	-0.0634	0.0357	0.0934	2.3798	3.7334	0.1542	0.1029	0.0287	0.2578	-0.0086	-0.0130	0.0054
6	2146.16	2312.34	3202.15	161.82	0.0859	-0.0163	0.0342	0.1270	3.1946	3.7224	0.1387	0.1533	0.0420	0.2578	-0.0086	-0.0148	0.0073
7	2150.81	2323.11	3016.46	148.32	0.1167	0.0308	0.0350	0.1632	3.2948	3.7331	0.1487	0.1235	0.0433	0.2620	-0.0041	-0.0112	0.0080
8	1841.43	2125.30	2942.56	159.96	0.1606	0.0896	0.0391	0.2205	3.0639	3.7993	0.1758	0.1025	0.0384	0.2697	0.0023	-0.0071	0.0109
9	1596.53	1890.03	2386.72	125.03	0.2524	0.1953	0.0456	0.3131	3.4677	3.8576	0.2255	0.1596	0.0299	0.2827	0.0218	0.0077	0.0145
10	973.86	1242.26	1571.08	63.69	0.8163	0.7198	0.0622	0.4730	4.9835	3.9048	0.3589	0.2456	-0.0182	0.3932	0.0824	0.0484	0.0315

^aThe donation of each variable for year t is suppressed in this table.
The statistics reported in this table are obtained from a sample that contains 103,965 firm-year observations for 10,480 distinct firms for the 38-year period (1973–2010).

The interesting variable is AQ, mean values of which decreases monotonically with the increase in AG ranks from AG rank of 1 to that of 6, then, increases monotonically with the increase in the AG ranks from AG ranks of 6 to that of 10. These results further support my H1A and H1B, the larger change (either upward with $AG > 0$ or downward with $AG < 0$) in sales growth rate is related to the poorer accruals quality (larger amount of AQ).¹⁸

The annual buy-and-hold return (Ret_f) for the year t monotonically increases with the increase in AG ranks, from annual returns of 0.0248 to of 0.4730 for portfolios with the AG ranks of 2-10. These results are consistent with the convention that the capital market rewards the high growth firms in the current year without controlling other factors. The mean value of annual returns for the portfolio with the lowest rank of AG is 0.0844, higher than that for portfolios with the AG ranks of 2-4, which may be due to the compensation of holding riskier assets.

The mean values of PB ratio (i.e., ratio of closing price to the book value of per share at the end of year) of two tails of AG ranks are larger ones, indicating the market prices firms with the more volatile in growth higher than firms with small change in sales growth (i.e., AG rank of 5). This may be because investors require more returns for holding riskier assets, partially supporting that AG is a risk proxy. The other side of the coin of compensating risk is that the market may overprice the lowest sales growth firms (the PB ratios for portfolios with AG ranks of 1 and 2 are 3.9789 and 8.7072, respectively). But these two portfolios have the highest portfolio abnormal returns in next year reported in Table 6. Therefore, the higher PB ratio in year t may not be due to the market mispricing. This result could be a puzzle for fundamental accounting researchers. Future research can investigate the low sales growth anomaly.

The business environment measures, operating cycle (OperCycle), and occurrence of negative earnings (NegEarn) are larger for the two tails of the AG ranks again. The longer the operating cycle and more frequent incidence of negative earnings are riskier of the firm. These results further suggest that AG is a risk proxy representing the change in the level of business risk of a firm.

The firms with the largest value of AG rank have the largest mean values of Leverage (0.3932), XFIN (0.0824), and Ddebt (0.0315), indicating that these firms are most active in external financing, and raise capital by borrowing. These results are consistent with convention that a firm in a high-growth state is more likely to expand its business by borrowing or issuing new/seasonal equity, which is inferred in Table 7.

The puzzle to us is that portfolio with the lowest AG rank, the slowest sales-growth-firm group, has the second highest external financing activity, and is the most likely to raise capital by issuing new/seasonal equity.

¹⁸Record, the larger (smaller) amount of AQ is the poorer (better) accruals quality.

103.6.3 Discussion

The abnormal sales growth (AG), the new risk proxy, proposed by this study represents a change in the level of operating risk, is associated with accruals quality and future returns. The results in this study have to be interpreted with caution because the sample firms used in this study tend to be large and profitable, compared to the Compustat population. The currently available accrual quality measures all suffer the possible measurement errors (Leone et al. 2006). I am aware that I may miss some explanatory variables of accruals quality as well. Further, some puzzles identified in this study could be future research.

103.7 Summary and Conclusion

I explore the possible relationship between the unexpected sales growth and accruals quality and future returns, inspired by Ohlson (2008) that analytically demonstrates a broader growth/risk equivalence concept and Penman (2010) that explains the uncertainty embedded in the sales growth in an accountant view. I argue that abnormal sales growth (AG) captures the uncertainty embedded in the sale growth, a type of operating (i.e., business) risk. AG is measured as the difference between sales growth rate in the current year and its benchmark, a weighted-average of sales growth rate over the past 3 years. The measure of AG suggests that a change in both directions of upward ($AG > 0$) and downward ($AG < 0$) reflects a change in business risk. This middle point, where AG is zero, is critical because it represents a firm's steady state of sales growth.

According to the value of AG, I divide my sample into a high- ($AG > 0$) and low-growth subsample ($AG < 0$). For the high-growth subsample ($AG > 0$), the larger the value of AG, the greater the uncertainty of sales growth (i.e., operating risk). For the low growth subsample ($AG < 0$), a larger value of AG indicates a lower operating risk because of less volatility in sales growth. The results for the high-growth subsample show that AG is positively correlated with the systematic risk (Beta) and the business risk proxies documented in Francis et al. (2005) (e.g., volatile of cash flows from operations, operating cycle, incidence of negative earnings over past 10 years). The converse is true for the low-growth subsample. Further, the results of the regressions of accruals quality (AQ) on AG show that the AG and AQ are positively (negatively) significantly associated, using the high-(low-) growth subsample. These results indicate that the accruals quality improves when the two tails of AG approach to zero. In other words, the accruals quality is the best when the firm is in the steady state.

Moreover, Francis et al. (2005) suggest that the accruals quality (AQ) includes two components: innate and discretionary components. The innate factors of accruals quality reflect the business environment of a firm. The results of this study extend the findings in Francis et al. (2005) by adding another explanatory factor to the accrual quality model. The contribution of the AG variable to estimation of accruals quality is highly significant: AG explains 9.7% (8.7%) of variation in accruals quality for high- (low-) growth subsample firms. The F test shows the

increments in R-squares are statistically significant at the 0.05 conventional levels. The magnitude of impact of AG on AQ is almost the same as that of volatility of sales (*STDSales*) on AQ in the regressions of accruals quality (AQ) on AG, *STDSales*, and other control variables. This indicates that AG captures different dimension of business risk other than volatility of sales and its impact on accruals quality is also economically meaningful compared to fundamental factors identified in the previous literature.

Finally, I examine the relationship between AG and future returns. The abnormal returns of the value-weighted portfolio with the lowest rank of AG in preceding year are the largest (18.48% annual abnormal returns on average). This result may be due to the market compensating stockholders of riskier assets. Exploring all possible detail reasons for this result is beyond the scope of this study, which could be a future research topic.

This study adds new evidence to accounting literature by demonstrating that the abnormal sales growth contains information uncertainty of future profitability estimated by expected future cash flows, which is reflected in the accruals quality, the proxy for information risk as Francis et al. (2005), and is associated with future returns. The empirical findings of this study confirm that accruals accounting is for risk (Penman 2016).

Acknowledgments Min (Shirley) Liu cordially thanks James A. Ohlson, Jeff Madura, Mark Kohlbeck, and workshop participants at Florida Atlantic University for their helpful comments on the early draft of this manuscript. The author also gratefully thanks the organizers of the 2017 Journal of Accounting, Auditing & Finance Conference, Cheng-Few Lee, and the participants of the research session of the 2021 AAA Spark and 2021 AAA Annual Meeting Webinar for their insightful suggestions on this manuscript. All remaining errors are my own.

Appendix

Accruals Quality

The calculation of total accruals quality (*AQ*), innate and discretionary component of accruals quality, follows the empirical procedures in Francis et al. (2004 and 2005), adopting the form that larger value of the *AQ* is a worse accruals quality. The accruals quality is measured on a firm- and year-specific basis, using the accounting data for rolling 10-year windows, $t-9 \dots t$. This procedure requires firms with time-series data. This requirement biases the sample toward surviving firms (larger and more successful firms). However, the advantage of using the firm as its own benchmark over long window outweighs the sample selection bias (Eason et al. 1992).

I measure accrual quality using modified Dechow and Dichev's (2002) model.

$TCA_{j,t} = \phi_0 + \phi_1 \cdot {}_jCFO_{j,t-1} + \phi_2 \cdot {}_jCFO_{j,t} + \phi_3 \cdot {}_jCFO_{j,t+1} + \phi_4 \cdot {}_j\Delta Rev_{j,t} + \phi_5 \cdot {}_jPPE_{j,t} + v_{j,t}$ (4) Where, $TCA_{j,t}$ = firm j's total current accruals in year t, calculated as $TCA_{j,t} = \Delta CA_{j,t} - \Delta CL_{j,t} - \Delta Cash_{j,t} + \Delta STDEBT_{j,t}$; $CFO_{j,t} = NIBE_{j,t} - TA_{j,t}$, firm j's cash flow from operations in year t; $NIBE_{j,t}$ = firm j's net income before extraordinary items

(Data #18) in year t ; $TA_{j,t} = \Delta CA_{j,t} - \Delta CL_{j,t} - \Delta Cash_{j,t} + \Delta STDEBT_{j,t} - DEPN_{j,t}$, firm j 's total accruals in year t ; $\Delta CA_{j,t}$ = firm j 's change in current assets (Data #4) between year $t-1$ and year t ; $\Delta CL_{j,t}$ = firm j 's change in current liabilities (Data #5) between year $t-1$ and year t ; $\Delta Cash_{j,t}$ = firm j 's change in cash (Data #1) between year $t-1$ and year t ; $\Delta STDEBT_{j,t}$ = firm j 's change in debt in current liabilities (Data #34) between year $t-1$ and year t ; $DEPN_{j,t}$ = firm j 's depreciation and amortization expense (Data #14) in year t ; $\Delta Rev_{j,t}$ = firm j 's change in revenues (Data #12) between year $t-1$ and year t ; $PPE_{j,t}$ = firm j 's gross value of PPE (Data #7) in year t . For each firm-year, I estimate equation (4) by industry (classified as Fama and French 1997) and year. These estimations yield firm- and year-specific residuals, $v_{j,t}$, which form the basis for the accrual quality metric, $AQ_{j,t} = \sigma(v_{j,t})$, equal to the standard deviation of firm j 's estimated residuals over year $t-4$ through t . The larger (small) the standard deviation of residuals is the poorer (better) the earnings quality.

Other Variables Used in Analyses

Sales = gross sales, scaled in millions.

AT = total assets, unit is million dollars.

Mcap = market value of the firm in million dollars.

IB = income before extraordinary items.

ROE = return on equity for year t .

SaleGrowth = sales growth in year t , computed as the difference of sales over time $t-1$ and time t , divided by sales in year $t-1$ ($\Delta Sales_{t-1,t} / Sales_{t-1}$).

AG = abnormal sales growth year t , computed as below.

$$AG_t = SaleGrowth_t - SalesGBechmark_t$$

$$SalesGBechmark_t = 0.3 \times SalesG_{t-3} + 0.3 \times SalesG_{t-2} + 0.4 \times SalesG_{t-1}$$

InnateAQ = innate component of accruals quality.

DisAQ = discretionary component of accruals quality.

Size = log value of total assets.

STDCFO = cash flow variability over the past 10 years.

STDSale = sales variability over the past 10 years.

OperCycle = operating cycle.

NegEarn = incidence of negative earnings over the past 10 years.

Leverage = total interest bearing debt to total assets.

PB = ratio of closing stock price per share at the end of fiscal year t to book value per share at the end of year t .

ROA = return on total assets for year t .

XFIN = external financing activities, computed as: $XFIN = DEquity + Ddebt$

DEquity = net cash received from the sale (and/or purchase) of common and preferred stock less cash dividend paid deflated by total assets

Ddebt = net cash received from the issuance (and/or reduction) of debt deflated by total assets [(DATA#111-DATA#114+DATA#301)/DATA#6]

Ret_f = 12-months buy-and-hold returns during the fiscal year $t-1$ to t , computed as:

$$Ret_f_{j,t} = \frac{P_{j,t} - P_{j,t-1}}{P_{j,t-1}} - 1$$

Where $P_{j,t}$ is firm j 's daily closing stock price at the end of fiscal year t , and $P_{j,t-1}$ is firm j 's daily closing stock price at the end of fiscal year $t-1$.

References

- Carhart, M.M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52 (1): 57–82.
- DeAngelo, L.E. 1986. Accounting numbers as market valuation substitutes: A study of management buyouts of public stockholders. *The Accounting Review* 61 (3): 400–420.
- Dechow, P. 1994. Accounting earnings and cash flows as measures of firm performance: The role of accounting accruals. *Journal of Accounting & Economics* 18: 3–42.
- Dechow, P.M., and I.D. Dichev. 2002. The quality of accruals and earnings: the role of accrual estimation errors. *The Accounting Review* 77: 35–59.
- Desai, H., S. Rajgopal, and M. Venkatachalam. 2004. Value-glamour and accruals mispricing one anomaly or two? *The Accounting Review* 79 (2): 355–385.
- Doukas, J.A., C. Kim, and C. Pantzalis. 2002. A test for the errors-in-expectations explanation of the value /glamour stock returns performance: evidence from analysts' forecasts. *Journal of Finance* 57 (5): 2143–2165.
- Eason, P., T. Harris, and J. Ohlson. 1992. Aggregate accounting earnings can explain most of security returns. *Journal of Accounting and Economics* 15: 119–142.
- Fama, E.F., and K. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.
- . 1997. Industry costs of equity. *Journal of Financial Economics* 43: 153–193.
- Fama, E.F., and J. MacBeth. 1973. Risk return and equilibrium: empirical tests. *Journal of Political Economy* 81: 607–636.
- Francis, J., R. LaFond, P.M. Olsson, and K. Schipper. 2004. Cost of equity and earnings attributes. *The Accounting Review* 79 (4): 967–1010.
- . 2005. The market pricing of accruals quality. *Journal of Accounting and Economics* 39: 295–327.
- Graham, B., and D. Dodd. 1934. *Security Analysis*. New York, NY: McGraw Hill.
- Jones, J.J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29: 193–228.
- Kothari, S.P., and J.B. Warner. 2007. *Handbook of Corporate Finance Volume 1*. Amsterdam: Elsevier North-Holland.
- Lakonishok, J., A. Shleifer, and R. Vishny. 1994. Contrarian investment, extrapolation, and risk. *Journal of Finance* 49 (5): 1541–1578.
- Leone, A.J., A. Kraft, and C. Wasley. 2006. An analysis of the theories and explanations offered for the mispricing of accruals and accrual components. *Journal of Accounting Research* 44: 297–339.
- Liu, M. 2021. What Information in Financial Statements Could be Used to Predict the Risk of Equity Investment? *Journal of Risk and Financial Management* 14 (8). Available at : <https://www.mdpi.com/1911-8074/14/8/365>
- McNichols, M.F. 2002. Discussion of the quality of accruals and earnings: the role of accrual estimation errors. *The Accounting Review* 77 (Supplement): 61–69.
- Ohlson, J. 2008. Risk, Growth, and Permanent Earnings. *Working Paper*. New York University.
- . 2014. Accruals: an overview. *China Journal of Accounting Research* 7 (2): 65–80.

- Penman, S. 2010. *Financial Statement Analysis and Security Analysis Valuation*. McGraw-Hill Irwin.
- . 2016. Valuation: Accounting for risk and expected returns. *Abacus* 52: 106–130.
- Petersen, M.A. 2009. Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies* 22: 435–480.
- Ohlson, J.A., and P. Bilinski. 2015. “Risk versus anomaly: A new methodology applied to accruals.” *The Accounting Review* 90 (5): 2057-2077.



Applications of Book-Tax Difference in Accounting and Finance Research

104

Nan-Ting Kuo and Shang-En Yu

Abstract

The difference between book income and taxable income or book-tax difference has received considerable interests from both the literature and policymakers. In this chapter we discuss the concept of book-tax difference and its empirical measures and we also review the literature exploring the information contained in book-tax differences. Prior studies consistently find that book-tax differences are informative about earnings management and tax avoidance behaviors. In addition, we also discuss the concept of book-tax conformity and its empirical measures, where it is a country-level factor in contrast to book-tax difference that is firm specific. The idea of book-tax conformity is highly related to that of book-tax differences because increasing book-tax conformity is a usually claimed means to curb the growing trend of book-tax differences as well as hinder the behind earnings management or tax avoidance behaviors. Prior literature suggests that increasing book-tax conformity may incur substantial costs in decreasing the informativeness of accounting earnings and it is debated about whether book-tax conformity can deter earnings management, while these question the potential benefits of increasing book-tax conformity.

Keywords

Book-Tax Difference · Book-Tax Conformity · Book Income · Taxable Income · Tax Rules · Accounting Standards

JEL Classification

M41 · K34

N.-T. Kuo (✉)

Shandong University of Technology, Business School, Zibo, Shandong, China

S.-E. Yu

Finance and Economics College, Jimei University, Jimei, China

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_107

2539

104.1 Introduction

In the United States, the growing divergence between book and taxable incomes has attracted much attention and analysis (Mills et al. 2002; Desai 2003; Hanlon and Shevlin 2005). This expanding divergence raises concerns about misleading financial or tax reporting because it signals that firms may increasingly overstate book income or understate tax liability. As the review made by Hanlon and Heitzman (2010) points out, the presence of book-tax differences raises considerable interests to the accounting and finance studies and it is linked to many important research areas such as the informational role of taxable income relative to book income or the measurement of corporate tax avoidance.

As the increasing trend of reported book-tax differences has received considerable interests in the literature, the first purpose of our paper is to review the literature regarding book-tax differences and its application in the accounting and finance studies. To this end we will clarify the concept of book-tax differences, elaborate commonly used empirical measures and their potential problems, and provide a literature review for studies exploring the effects of book-tax differences. Our literature review focuses on the information content of book-tax differences as the presence of book-tax differences is an indicator for aggressive accounting or tax reporting practices.

In addition, we will discuss book-tax conformity, where we emphasize that it is a country-level measure in contrast to book-tax difference that is a firm-level measure. Book-tax conformity is closely related to the literature of book-tax differences because the former is a plausible strategy to mitigate the increasing growth trend in firm-level book-tax differences as well as the accompanied concerns about earnings manipulation and tax avoidance. We will discuss common empirical measures of book-tax conformity and review the related literature. Following Hanlon and Heitzman (2010), our literature review focuses on the effects of increasing book-tax conformity on decreasing the informativeness of accounting earnings numbers as this is a major debate regarding whether it deserves to increase the degree of book-tax conformity.

This chapter is organized as follows. In Sect. 104.2 we discuss the concept of book-tax differences and its empirical measures. Section 104.3 reviews the literature exploring the effects of book-tax differences. Section 104.4 introduces the concept and empirical measures of country-level book-tax conformity, which is closely related to the idea of firm-level book-tax differences. Section 104.5 reviews the literature exploring the effects of book-tax conformity. Concluding remarks appear in Sect. 104.6.

104.2 The Concept and Empirical Measures of Book-Tax Differences

Conceptually, book-tax differences arise from the divergences between accounting income (i.e., book income) and taxable income and they are driven by various factors. The fundamental of these differences is that accounting income and taxable

income are measured with two systems with quite different objectives that in turn lead to different rules. The reasons for their differences are discussed as follows, where our discussions are abstracted from Hanlon and Heitzman (2010).

In theory, the measurement of accounting income follows the conceptual framework of financial accounting standards (GAAP), where the main objective is to capture the underlying economics of transactions and provide relevant information for financial reporting users such as equity investors or contracting parties. By contrast, taxable income is measured based on tax rules, while tax rules are enacted for a more political purpose in that they are purported to raise fiscal revenues for governments, direct certain private sector activities such as encouraging firms to engage in R&D and innovation activities, or stimulate the economy through for example cutting personal income taxes to enhance consumption.

Moreover, while the calculation of accounting income is accrual based, the calculation of taxable income is basically a mix of cash basis and accrual basis, which does not allow recognizing estimated expenses before the realization of cash payments or deferring the recognition of unearned revenues. In particular, the location where the earnings made is crucial because it is the jurisdiction of the location that can levy tax on the earnings. By contrast, the location where the earnings are made is irrelevant under financial accounting because financial accounting standards require that all earnings from controlled entities are aggregated into a consolidated financial statement.

Furthermore, managers' aggressive reporting for accounting or tax purposes is also a likely source of the book-tax differences, where it may result from aggressive financial reporting due to upward earnings management or from aggressive tax reporting due to tax avoidance. Specifically, when managers engage in upward earnings management, they will record a book-tax difference in the financial statements if they choose to report unmanaged taxable income. On the other hand, when managers engage in tax avoidance and take actions to lower reported taxable income, they will also record a book-tax difference if they choose to report unmanaged accounting income.

Note that reporting a book-tax difference is a less costly method for either aggressive financial reporting or aggressive tax reporting. For aggressive financial reporting, book-tax differences provide relevant information to market participants about the quality of reported earnings in the sense that recoding book-tax differences hurts investor perceptions about the credibility of reported earnings. To avoid reducing perceived earnings quality, managers can choose to report higher amount of taxable income and paying taxes on the inflated earnings they manipulate, but this leads to higher tax payments, suggesting that recording book-tax differences saves taxes and therefore it is a less costly method.

For aggressive tax reporting, reporting book-tax differences is likely to attract the attentions from the taxing authority and it may be used as an indicator of tax avoidance, which increases the likelihood of detection for the understatement of taxable income by the taxing authority. To avoid detection by the taxing authority, managers can choose to report lower accounting income to conform to the understated taxable income, but this leads to higher reporting costs due to the lower reported earnings so managers may still choose to record book-tax differences.

Accordingly, there exists a trade-off between recoding book-tax differences and paying higher taxes or bearing higher reporting costs when manager attempt to engage in aggressive tax or accounting reporting.

The above discussions suggest that accounting income and taxable income are linked to some extent but in most situations they are not conformed for various transactions, leading to the call for book-tax conformity that we will discuss in subsequent sections. In most cases, all of these factors operate concurrently to determine the level of book-tax differences a firm reports. If book income is calculated with the same rules as taxable income, then book-tax differences will not present. However, because accounting rules divergent from tax rules in that the former focuses on measuring economic performance while the latter focuses on levying taxes, the presence of book-tax differences is quite common. In this regard, it is well recognized in the literature that there are many possible determinants of book-tax differences and prior studies usually intend to develop models or identify settings to study a particular determinant.

To conduct the research, prior studies typically partition the book-tax difference measures based on the underlying causes of the relations explored, but doing so raises an important issue that this stream of literature employs several measures of book-tax differences because it is usually not clear about the reasoning for why a particular book-tax difference measure is used for the research question explored. We clarify this issue with following reviews about the commonly used measures of book-tax differences, where our classifications and explanations closely follow Evers et al. (2016) by condensing their arguments.

Recall that a firm's reported book-tax difference is calculated as the difference between its reported book income and taxable income, where book income is the income from financial statements and taxable income is the income used to determine tax liability. Because data for taxable income is unavailable, it is usually inferred by grossing up current tax expenses with the statutory corporate income tax rate. In this regard, the basic measure of book-tax differences (denoted as total BTD) is as presented in Eq. (1).

$$\text{Total BTD} = \text{Pre-tax book income} - \text{Estimated taxable income} \quad (1)$$

$$\text{Estimated taxable income} = \text{Current tax expenses} / \text{Statutory tax rate} \quad (2)$$

Equation (1) presents the basic definition of book-tax differences as it captures both the temporary and the permanent components of book-tax differences. Equation (2) shows how to estimate taxable income. Since tax return data are usually unavailable, taxable income must be estimated with current tax expenses retrieved from financial statements (Manzon and Plesko 2002).

Although total BTD is intuitively appealing due to its straightforward computation, it is criticized as being subject to serious measurement errors arising from, for instance, different consolidation rules for book and tax purposes, tax credits, foreign operations, or loss firms (Hanlon 2003; Kvaal and Nobes 2012; Evers et al. 2016). Hence, total BTD is rarely used as the main proxy in the research design, and it must

be complemented by other measures of book-tax differences to ensure the robustness of empirical results.

As summarized in Evers et al. (2016), some studies (e.g., Hanlon et al. 2012) separate total BTD into temporary BTD and permanent BTD, which are calculated based on following Eqs. (3) and (4) respectively.

$$\text{Temporary BTD} = \text{Deferred tax expenses} / \text{Statutory tax rate} \quad (3)$$

$$\text{Permanent BTD} = \text{Total BTD} - \text{Temporary BTD} \quad (4)$$

Temporary BTD arises due to the timing difference in the recognition between book income and taxable income with regard to specific revenue or expense items. Temporary BTD can be estimated by grossing up deferred tax expenses with the statutory tax rate (Moore 2012) as the timing difference is reconciled through the recognition of deferred tax expenses. A specific example is that many firms use different depreciation methods to calculate book income and taxable income, where the difference between these two incomes is a timing difference and it is temporary in nature as it will reverse in the future. Accordingly, temporary BTD is regarded as containing information about non-tax accruals such as depreciation (Hanlon and Heitzman 2010).

Permanent BTD arises from the difference between book and taxable income that is not a timing difference and thereby will not reverse in the future, where it is determined with the difference between total BTD and temporary BTD (Wilson 2009; Lisowsky et al. 2013). Permanent BTD is used as an indicator of tax shelters in the literature (Plesko 2004; Frank et al. 2009; Shevlin 2002) because it represents a decrease in taxable income without affecting book income, while this feature suggests that the presence of permanent BTD is indicative of aggressive tax reporting. Consistent with this idea, Wilson (2009) provides empirical evidence that the construction of most tax shelters leads to permanent BTD.

To more precisely capture tax reporting aggressiveness, Frank et al. (2009) develop a model as following Eq. (5) to measure discretionary permanent differences (denoted as *DTAX*), where *DTAX* is measured with the regression residuals. By considering factors reflecting the non-discretionary portion of permanent BTD, the residuals represent the portion of permanent BTD that removes the effects of determinants unrelated to tax aggressiveness so they are supposed to capture intentional tax aggressiveness.

$$\text{PermanentBTD} = \alpha + \beta * \text{Non-discretionary items} + \gamma * \text{Other statutory adjustments} + \varepsilon \quad (5)$$

In Eq. (5), *DTAX* is estimated with the residual term ε . The model of Frank et al. (2009) is based on the idea that the construction of tax shelters often leads to the presence of permanent BTD. Accordingly, the development of their model is intended to exclude the effects of tax planning resulted from temporary differences. In addition, permanent BTD may reflect items not considered as related to aggressive

tax reporting such as changes in the tax cushions, changes in valuation allowance, goodwill and other intangible assets or tax credits. Hence, *DTAX* is estimated by regressing permanent BTD on items known to induce non-discretionary permanent differences and statutory adjustments that are likely unrelated to tax aggressiveness, where the resultant residuals remove the effects of determinants unrelated to tax aggressiveness and thus they are supposed to capture the intentional tax aggressiveness or the discretionary portion of permanent BTD.

Similarly, Desai and Dharmapala (2006) develop a model as Eq. (6) to measure discretionary total BTD, which is intended to capture the portion of total BTD that is most likely to represent tax aggressiveness, where the residual term ε represents discretionary total BTD. Specifically, the first step of their model is to calculate total BTD based on Manzon and Plesko (2002) and the second step is regressing total BTD on total accruals.

$$\text{Total BTD} = \beta * \text{Total accruals} + \text{Firm fixed effects} + \varepsilon \quad (6)$$

The behind logic for Eq. (6) is that total BTD can be attributable to earnings management and tax avoidance. To the extent that total accruals capture the effects of earnings management, the residual from the regression represent the portion of total BTD that cannot be explained by earnings management and thereby it can represent the effects of tax avoidance. Econometrically speaking, because earnings management is reflected on accruals, the control for total accruals removes the effects of earnings management from total BTD, which leads residuals to capture the remaining effects of total BTD. In this regard, discretionary BTD estimated from Eq. (6) represents the part of total BTD that can be attributable to tax avoidance.

The final measure of book-tax differences is developed by Wilson (2009) who estimates BTD by removing the effects of tax sheltering, where the ultimate aim is determining BTD of tax shelter firms and non-shelter ones. Using a sample of firms that were disclosed to engage in tax shelter by the US Treasury or by the Press, his approach first determines the tax savings from the notes of these firms' financial statements and then determines the tax benefits by grossing up the identified tax savings with applicable statutory income tax rate. His measure of book-tax differences is obtained by removing those tax benefits from total BTD.

It is noteworthy that all measures of book-tax differences require the estimate of taxable income as an input so the measurement errors from estimated taxable income are a common concern underlying all measures of book-tax differences. Thus, it deserves further discussions about the problems of estimating taxable income with Eq. (2).

Specifically, using current tax expenses to infer taxable income is a common approach employed by numerous prior studies (e.g., Gleason and Mills 2002; Lev and Nissim 2004). The underlying logic behind this approach is that since current tax expenses contain the taxes levied on the taxable income of current year, an estimate of taxable income can be inferred by dividing current tax expenses by the tax rate. This is particularly the case because although taxable income is unobservable, current tax expenses are usually disclosed in the notes of financial statements.

However, this approach leads to a biased estimate of taxable income for a number of reasons. First, from the perspective of income statement presentation, tax expense and taxable income do not exactly match because tax expense reflects only items regarding profit and loss from continuing operations and it excludes taxes related to other comprehensive income or taxes related to equity and discontinued operations. By contrast, tax return does not present similar partition of these items. Because current tax expense is one of the two components (the other is deferred tax expense) constructing tax expense, the mismatch of presentations between income statement and tax return naturally induces measurement error to estimating taxable income from current tax expenses (Kvaal and Nobes 2012).

Second, this approach holds only with a single tax jurisdiction (Hanlon 2003). For a multinational company that is taxed by different jurisdictions, it is necessary to separately estimate domestic and foreign current tax expenses and grossing up them with domestic and foreign tax rates respectively. However, the information disclosed in financial statements is not always sufficient to enable this approach as the disclosure of foreign tax expenses is usually absent. This is particularly troublesome if a company has operations over several countries so that its foreign tax expenses are an aggregation of multiple tax rates across different jurisdictions. Consequently, for a taxable entity with operations across several jurisdictions, even manually collected data from tax disclosures may not ensure an accurate estimate of taxable income. Relatedly, special tax regimes from such as extractive industries may confound the relation presented in Eq. (2), while this suggests that Eq. (2) only holds when current tax expenses is a function of statutory tax rate.

Third, taxation is asymmetric in that governments only tax profits but do not share losses, while this feature precludes negative taxes. Therefore, for a negative taxable income, current tax expenses should be zero and the relation in Eq. (2) will not hold. As suggested by Atwood et al. (2010), a possible solution to address this issue is to exclude all sample firms with negative pre-tax book income or negative current tax expenses, but this solution unavoidably omits a large proportion of sample and it may thus not only confound the reliability of empirical models but also affects the interpretation of the results. A related issue is that net operating loss carryforwards and carrybacks may lead to negative current tax expenses, while this complicates the application of Eq. (2) because offsetting tax losses to taxable income makes current tax expenses deviate from true taxable income. As suggested by Manzon and Plesko (2002), a possible solution is to subtract the changes in the balances of tax loss carryforwards from Eq. (2). However, a primary limitation of this approach is data availability for loss carryforwards since their disclosures are not always mandatory and even if mandatory (e.g., as required by IFRS or US GAAP), the information disclosed is typically not standardized and this enhances the difficulty in collecting this information.

Fourth, the difference in the consolidation rules between financial and tax reporting may also induce noises to using current tax expenses as a proxy for taxable income. This is exactly the case for the US environment (Manzon and Plesko 2002) and even more so for comparisons across countries due to the substantial differences in consolidation rules across countries.

Finally, adjustments of prior years' taxable income or reverses of tax cushions for contingent tax liabilities make current tax expenses deviate from the function of current year's taxable income, which further complicates the application of Eq. (1).

In summary, although the relation presented in Eq. (1) is straightforward, it is subject to a number of flaws. When applying book-tax differences in the empirical tests, we need to scale them by such as total assets or market value of equity. The estimated book-tax differences may have positive or negative values.

104.3 Review of Literature Related to Book-Tax Differences

Prior literature provides ample evidence supporting that both income tax expenses and reported book-tax differences contain useful information to the market, where such information derives from the idea that taxable income is an alternative performance measure or a benchmark measure relative to financial accounting income. To date, the evidence suggests that book-tax differences contain information relevant to evaluate potential earnings management or tax avoidance behaviors, where book-tax difference measures are specifically designated to investigate aggressiveness in either or both financial and tax reporting. The results of this stream of literature are applicable to both accounting and finance studies as the information role of taxable income is a common topic in both studies, while such information effect is distinguished from the effect of tax avoidance. Because this research requires knowledge in both tax rules and accounting standards, it is often conducted by tax researchers.

Specifically, the focus of prior literature is on the link between book-tax differences and aggressive reporting practices, where higher book-tax differences are not only associated with more earnings management or lower financial reporting quality (e.g., Badertscher et al. 2009; Frank and Rego 2006; Hanlon 2005; Phillips et al. 2003, 2004; Tang and Firth 2011) but also associated with more tax avoidance (e.g., Desai 2003; Frank et al. 2009; Mills 1998; Tang and Firth 2011; Wilson 2009).

The idea that book-tax differences contain the information about earnings quality is well recognized (e.g., Penman 2001). The behind reasoning for this idea is that accounting accruals allow more discretion than do tax rules and thereby temporary book-tax differences are informative about the discretion embedded in non-tax accounting accruals such as bad debt expenses, warranty expenses or deferred revenues. This reasoning is not applicable to permanent book-tax differences because they are not driven by the process in recognizing accounting accruals. It is also suspected that firms reporting book income higher than taxable income are likely to manipulate book income upward, where such suspicion arises from the Enron and WorldCom scandals. However, it is difficult to judge manipulation based on the sign of book-tax differences because while positive book-tax differences signal upward earnings management, negative book-tax differences (i.e., book income is lower than taxable income) could also lead to earnings quality concerns if firms take a "big bath" or use "cookie-jar" reserves, given that related write-offs or expense accruals are often not deductible for tax purposes.

To examine the information contained in book-tax differences for earnings quality, prior studies have adopted various empirical designs and the overall results are quite consistent, where book-tax differences are informative about the quality of financial accounting earnings. For example, Joos et al. (2000) find that firms reporting large book-tax differences (in absolute value) have weaker earnings-return relations. Phillips et al. (2003) find that firm-year observations with a small earnings increase tend to have larger deferred tax expenses, while this suggests that deferred tax expenses (or temporary book-tax differences) contain information about pre-tax earnings management (also refer to Phillips et al. 2004). Hanlon (2005) presents evidence that book-tax differences are informative about the persistence of pre-tax earnings and pre-tax accruals, while the market understands this information implication. Lev and Nissim (2004) document a positive relation between future earnings growth and the ranked ratio of taxable income over book income and this ratio can predict stock returns, suggesting that equity investors may have difficulty in processing the information contained in book-tax differences from a pricing perspective. Based on the perspective of credit ratings agencies, Ayers et al. (2010) find that large changes in book-tax differences seem to be associated with more negative revisions in credit ratings.

Extending this stream of literature, Chi et al. (2014) provide evidence consistent with market mispricing of book-tax differences. Weber (2009) finds that analysts as sophisticated financial statement users do not fully understand the information contained in book-tax differences when making earnings forecasts. Dhaliwal et al. (2017) find a positive association between the variation in estimated taxable income and cost of equity capital. Hanlon et al. (2012) and Kuo and Lee (2016) find that larger book-tax differences are associated with higher audit fees, suggesting that auditors also use the information of book-tax differences to plan their works. By separating total book-tax differences into temporary and permanent portions, Moore and Xu (2018) report that temporary book-tax differences, either levels or variability, are consistently positively associated with costs of private debt, whereas permanent book-tax differences are not.

On the other hand, there is also evidence suggesting that book-tax differences contain information about tax avoidance behaviors. For example, Mills (1998) reports that firms with large book-tax differences are more likely to be audited by the IRS and thereby have larger audit adjustments. Wilson (2009) finds that firms accused of engaging in tax shelter tend to have larger book-tax differences. Further, recent studies provide mixed evidence on the effects of borrowers' book-tax differences on private loan costs. Hasan et al. (2014) find that tax avoidance measured with book-tax differences is positively associated with covenant requirements and interest rate spreads, particularly for firms with higher discretionary accruals. By contrast, Kim et al. (2010) find a negative association between aggressive tax avoidance and private loan costs. Desai (2003) argues that the increasing divergence between book and tax income is evidence of aggressive tax reporting. However, Shevlin (2002) notes that it must be cautious by drawing inferences about the levels or trends of tax avoidance based on book-tax differences.

The above studies suggest that book-tax differences capture at least some element of tax avoidance. Of course, book-tax differences by definition only capture non-conforming tax avoidance, while this limits using them as a measure to compare tax avoidance activities across firms with different levels of importance on financial accounting earnings.

104.4 The Concept and Empirical Measures of Book-Tax Conformity

It has been an ongoing debate in the literature and among policymakers about whether book and tax income measures should conform to one common measure (i.e., increase the degree of book-tax conformity). This debate arises because there is a growing divergence between book and taxable incomes (Mills et al. 2002; Desai 2003; Hanlon and Shevlin 2005), while such divergence raises concerns about misleading financial or aggressive tax reporting because it signals that firms may increasingly overstate book income or understate tax liability, leading to more firm-level reported book-tax differences as elaborated in previous sections.

A potential solution to mitigate these concerns is increasing the conformity between book and taxable incomes because increased book-tax conformity enhances the cost for firms to simultaneously manage taxable income downward and book income upward (Yin 2001; Desai 2005). Accordingly, book-tax conformity has emerged in policy circles as a means of reducing the perceived ability of firms to manage earnings and evade taxes.

Proponents of increased book-tax conformity (e.g., Yin 2001; Desai 2003, 2005; Whitaker 2005; Shaviro 2009) argue that a higher degree of book-tax conformity constrains managers' discretion and incentives to engage in aggressive financial and tax reporting because increased book-tax conformity forces firms to trade off the desire to report high book income numbers to shareholders and the desire to report low income numbers to tax authorities, resulting in improved financial reporting quality and lesser tax avoidance. In addition, increased book-tax conformity also leads to lower compliance costs. For example, Kuo and Lee (2016) find that a higher degree of book-tax conformity is associated with lower fees that firms pay for their auditors.

However, opponents (Hanlon et al. 2005, 2008; Hanlon and Shevlin 2005; McClelland and Mills 2007) question the claimed benefits from such forced trade-off because the implementation of book-tax conformity would not be as simple as the proponents argue and it is also uncertain about the unintended consequences due to behavioral responses on the part of firms. More importantly, increased book-tax conformity incurs a substantial cost – a decrease in the information content contained in accounting earnings numbers – because the objectives of accounting and tax reporting are divergent and enforced conformity will lead accounting earnings numbers to deviate from its objective of providing useful information to guiding investment decisions.

We then review the empirical measures of book-tax conformity. Before proceeding, we first clarify that book-tax conformity pertains to the extent to which book and taxable incomes are conformed so it represents a country-level measure, which is in contrast to firm-level book-tax difference that represents a gap between a firm's accounting income and taxable income that are reported on financial statements and tax returns respectively. Prior studies exploring the effects of book-tax conformity mostly use a cross-country sample because this research design can exploit the variations of book-tax conformity across countries. This is a natural design since book-tax conformity is a country-level concept with the levels varying across jurisdictions (Atwood et al. 2012).

When measuring book-tax conformity, early studies (Ali and Hwang 2000; Guenther and Young 2000; Hung 2001; Burgstahler et al. 2006) rely on a dummy variable that simply categorizes the contemplated countries into high and low conformity. This binary categorization is derived based on the similarity between tax laws and accounting standards in measuring income, which represents the perceived extent that accounting provisions of the tax law conform to financial accounting standards (Watrin et al. 2014). Hence, this binary measure is rather subjective (Tang 2015) and it also raises a serious econometric issue from correlated omitted country-level institutional factors.

In contrast to this subjective binary measure, Atwood et al. (2010) develop a model to estimate the level of required book-tax conformity across country-year, which measures "the flexibility that a firm has to report taxable income that is different from pre-tax book income." Their model is presented as following Eq. (7).

$$CTE = \alpha + \beta_1 * PTBI + \beta_2 * ForPTBI + \beta_3 * DIV + \varepsilon \quad (7)$$

where *CTE* is current tax expense, including both domestic and foreign current tax expenses; *PTBI* is pre-tax book income; *ForPTBI* is the estimated foreign pre-tax book income (foreign tax expense/total tax expense $\times$ *PTBI*); *DIV* is dividends; *CTE*, *PTBI*, *ForPTBI*, and *DIV* are scaled by total assets.

The book-tax conformity measure is calculated as the scaled ranking of the root mean squared errors (RMSEs) obtained from country-year estimates of Eq. (7). Because RMSE in Eq. (7) represents the standard error of *CTE* for a given level of *PTBI* not explained by the model, it provides an indication of the overall amount of discretion that managers have to report different book income and taxable income. Consequently, a higher RMSE corresponds to lower book-tax conformity.

Atwood et al.'s (2010) measure is widely used in the literature. However, a potential concern for their measure is that its estimate is based on the information retrieved from consolidated financial statements and this may confound the validity of using current tax expenses to represent taxable income.

To address the issues due to consolidated financial statements, Watrin et al. (2014) develop a book-tax conformity measure based on the relation between single financial statements and tax accounts, where they justify this approach by claiming that the determination of taxable income is according to single financial statements in European countries featured with high book-tax conformity. Watrin et al. derive their

measure by first calculating permanent BTD at the single entity level and then assigning a rank to each country-year based on the country-year mean of the absolute values of permanent BTD, such that countries with higher rankings have lower aggregated BTDs and thereby have higher book-tax conformity. .

Based on similar approach as Atwood et al. (2010), Tang (2015) develops another measure by defining required book-tax conformity as the variation in total BTD “that cannot be explained by opportunistic book and tax reporting for firms in a given country and year.” Specifically, Tang’s measure is targeted at disentangling total BTD into the portion arising from the innate differences between accounting and tax rules and the portion arising from opportunistic book and tax reporting. To this end, she first regresses total BTD on the proxies for earnings management and tax avoidance and their interaction term, where earnings management is proxied by discretionary accruals and tax avoidance is proxied by the difference between the statutory tax rate and the effective tax rate. She then uses the RMSEs from the regression to measure the extent to which book and taxable incomes diverge due to the innate differences between accounting and tax rules, thereby reflecting the degree of required book-tax conformity.

The measures developed by Atwood et al. (2010), Watrin et al. (2014), and Tang (2015) are similar in that all of them use pre-tax book income and taxable income to gauge the degree of book-tax conformity. However, this raises a concern that while a higher required book-tax conformity leads firms to report lesser book-tax differences, the opposite is not necessarily true and we cannot conclude that there is a complete book-tax conformity just because we observe that pre-tax income equals taxable income. Even in an environment where accounting and tax rules are quite divergent and the degree of book-tax conformity is low, the different components of income may offset each other so that pre-tax income coincidentally equals to taxable income as if these two incomes were conformed. Nevertheless, despite their imperfection these measures are still important to the literature as they provide a numerical measure that facilitates subsequent studies to investigate the effects of book-tax conformity.

Indeed, although there are many divergences between accounting and tax rules in measuring income, these two systems are aligned at some level. While these two income amounts often diverge to a large extent in a dual reporting system, there is still basic alignment in the fundamental accounts of such as sales or cost of goods sold. Consistent with this notion, there is a stream of literature exploring whether tax cuts drive earnings management for the purpose of saving taxes, where firms shift income from high- to low-taxed years and the evidence supports significant shifting in financial accounting earnings (Scholes et al. 1992; Guenther 1994; Maydew 1997).

104.5 Review of Literature Related to Book-Tax Conformity

As argued by Hanlon and Heitzman (2010), it is difficult to examine what would happen if book-tax conformity were adopted because an idea research design requires a switch from a book-tax conformity system to a non-conformed system

(or vice versa). Absent such as natural experiment, most of the research has primarily focused on a testable item: whether increased book-tax conformity decreases the information content of accounting earnings numbers. Despite the weakness of research designs, most prior studies document a substantial loss of information in accounting earnings.

Specifically, while accounting standards allow managerial discretion that could decrease the informativeness of earnings due to managers' opportunistic reporting incentives, it is possible that managers use the discretion to convey private information to the market rather than use it to mislead investors (e.g., Dechow 1994; Subramanyam 1996; Beaver and Engel 1996). That is, managers use at least some of their discretion on the recognition of accounting numbers to signal value relevant information to market participants. Consequently, if managers were required to report earnings based on tax rules as in an environment with high book-tax conformity, they would lose at least some of their ability to signal value relevant information through reporting accounting numbers. Furthermore, increased book-tax conformity would increase firms' incentives to report lower accounting earnings numbers to avoid taxes (e.g., the use of LIFO for inventory accounting even though it lowers reported book income), where this adds bias to the earnings numbers as a measure of economic income.

Consistent with this line of reasoning, Hanlon et al. (2005) provide evidence that taxable income estimated based on accounting rules in the U.S. is about 50% less informative to equity investors than book income, while this suggests that the loss of information to investors is likely significant if book income were recognized on the basis of tax rules. Using a more powerful research design, Hanlon et al. (2008) explores a sample of US firms required to convert from cash-basis tax accounting to accrual-basis tax accounting with no change in their financial reporting methods, where these firms used accrual basis accounting before and after the tax law change for book purposes. They find that relative to control firms that always used the accrual basis for tax accounting, converting firms have less informative earnings after conversion, while this finding suggests that requiring firms to pay taxes on book income leads to information loss on earnings. They further argue that "noise in earnings could increase with conformity because of managers' inability (due to the tax cost of doing so) to convey private information useful to external stakeholders through earnings. If managers are constrained in relaying this private information via earnings, noise in earnings will increase, reducing the ERC (earnings response coefficient). Conversely, if book-tax conformity causes firms to report more accurately, as its proponents suggest, then noise will decrease and the ERC will increase. Our empirical results are consistent with increased book-tax conformity increasing the amount of noise in earnings, reducing its informativeness."

In addition to explore US firms, another stream of literature uses an international setting to explore the effects of book-tax conformity on earnings informativeness. By comparing the effects of country-specific factors on earnings informativeness, Ali and Hwang (2000) find that earnings are less informative in countries with higher book-tax conformity. Guenther and Young (2000) find that earnings are more closely linked to economic activity in countries such as the USA or the UK because in these

countries accounting standards are more independent from the tax standards, allowing managers more discretion to convey private information and thereby leading earnings to better reflect economic performance. Atwood et al. (2010) explore whether increased book-tax conformity affects the persistence of earnings and the ability of earnings to predict future cash flows. They find that both earnings persistence and the ability of earnings to predict future cash flows are low when book-tax conformity is high, where their findings suggest that increased book-tax conformity likely reduces earnings quality.

Building on the insight that book-tax conformity induces information loss of earnings, Blaylock et al. (2017) argue that such information loss is more for equity investors than for debt investors and they further document that increased book-tax conformity leads firms to rely more on debt capitals, suggesting that book-tax conformity is also a finance topic as it affects capital structure. Atwood et al. (2012) report that increased book-tax conformity reduces the incentives for firms to engage in tax avoidance.

Some studies explore how book-tax conformity affects earnings management. Using an international setting, Leuz et al. (2003) employ a dummy variable indicating whether a country has high or low book-tax conformity and find an insignificant association between book-tax conformity and earnings management. Burgstahler et al. (2006) use a similar setting and they also find no significant association between book-tax conformity and earnings management for public firms, while this association is positive for private firms. Using the book-tax conformity measure of Atwood et al. (2010) and similar earnings management measures as Leuz et al. (2003), Burgstahler et al. (2006), and Blaylock et al. (2015) document a robust positive association between book-tax conformity and earnings management in an international setting. However, Tang (2015) uses her book-tax conformity measure and finds that a higher level of book-tax conformity is associated with lower earnings management. The results of these studies suggest that there is still no consensus about whether increased book-tax conformity can deter earnings management.

Overall, prior evidence suggests that increasing book-tax conformity decreases earnings informativeness because it reduces the ability of managers to use their discretion to signal value relevant information given that tax incentives confound managers' intent to recognize income in a timely manner and induce them to report a book income number targeted at tax planning rather than measuring economic performance. However, as Hanlon and Heitzman (2010) point out, we know very little about how book-tax conformity affects other corporate decisions.

104.6 Conclusion

Our entry discusses the concept and empirical measures of book-tax differences. Although there are a variety of book-tax difference measures, all of them are based on a key element: estimated taxable income. Because tax return data are usually unavailable, taxable income must be estimated and this raises concerns about the measurement errors as well as the validity of the resultant book-tax

difference measures. We also review the literature exploring the effects of book-tax differences and the literature reviewed consistently suggests that book-tax differences contain useful information to market participants, where this information derives from the idea that taxable income is an alternative performance measure relative to book income. In this regard, prior studies suggest that book-tax differences contain information relevant to evaluate potential earnings management or tax avoidance behaviors. In addition, we also discuss the concept and empirical measures of book-tax conformity, given that increasing the degree of book income aligning to taxable income is a usually claimed means to curb the growing trend of book-tax differences reported by firms as well as the behind aggressive financial or tax reporting practices. The literature we reviewed about book-tax conformity suggests that increasing book-tax conformity will decrease the informativeness of accounting earnings numbers. Given that there is still debated about whether increasing book-tax conformity can deter earnings management as its proponents claimed, the information loss induced by book-tax conformity might incur substantial costs that more than offset the benefits of conformity from such as reducing compliance costs. Overall, as Hanlon and Heitzman (2010) suggest, we need more evidence to better understand the influences of increasing book-tax conformity on corporate decisions.

References

- Ali, A., and L. Hwang. 2000. Country-specific factors related to financial reporting and the value relevance of accounting data. *Journal of Accounting Research* 38: 1–21.
- Atwood, T.J., M.S. Drake, and L.A. Myers. 2010. Book-tax conformity, earnings persistence and the association between earnings and future cash flows. *Journal of Accounting and Economics* 50: 111–125.
- Atwood, T., M. Drake, J. Myers, and L. Myers. 2012. Home country tax system characteristics and corporate tax avoidance: International evidence. *The Accounting Review* 87: 1831–1860.
- Ayers, B., S. Laplante, and S. McGuire. 2010. Credit ratings and taxes: The effect of book/tax differences on ratings changes. *Contemporary Accounting Research* 27 (2): 359–402.
- Badertscher, B., J. Phillips, M. Pincus, and S. Rego. 2009. Earnings management strategies and the trade-off between tax benefits and detection risk: To conform or not to conform? *The Accounting Review* 84 (1): 63–97.
- Beaver, B., and E. Engel. 1996. Discretionary behavior with respect to allowance for loan losses and the behavior of security prices. *Journal of Accounting and Economics* 22: 177–206.
- Blaylock, B., B.F. Gaertner, and T. Shevlin. 2015. The association between book-tax conformity and earnings management. *Review of Accounting Studies* 20 (1): 141–172.
- . 2017. Book-tax conformity and capital structure. *Review of Accounting Studies* 22: 903–932.
- Burgstahler, D., L. Hail, and C. Leuz. 2006. The importance of reporting incentives: Earnings management in European private and public firms. *The Accounting Review* 81 (5): 983–1016.
- Chi, S., M. Pincus, and S.H. Teoh. 2014. Mispricing of book-tax differences and the trading behavior of short sellers and insiders. *The Accounting Review* 89 (2): 511–543.
- Dechow, P. 1994. Accounting earnings and cash flows as measures of performance: The role of accounting accruals. *Journal of Accounting and Economics* 18: 3–42.
- Desai, M. 2003. The divergence between book income and tax income. In *Tax policy and the economy*, ed. J. Poterba, vol. 17, 169–206. Cambridge: MIT Press.

- . 2005. The degradation of reported corporate profits. *Journal of Economic Perspectives* 19: 171–192.
- Desai, M., and D. Dharmapala. 2006. Corporate tax avoidance and high-powered incentives. *Journal of Financial Economics* 79 (1): 145–179.
- Dhaliwal, D., H.S. Lee, M. Pincus, and L. Steele. 2017. Taxable income and firm risk. *Journal of the American Taxation Association* 39 (1): 1–24.
- Evers, M.T., I. Meier, and K. Nicolay. 2016. The implications of book-tax differences: A meta-analysis. ZEW – Centre for European Economic Research Discussion Paper No. 17-003.
- Frank, M.M., and S. Rego. 2006. Do managers use the valuation allowance account to manage earnings around certain earnings targets? *Journal of the American Taxation Association* 28 (1): 43–65.
- Frank, M.M., L.J. Lynch, and S.O. Rego. 2009. Tax reporting aggressiveness and its relation to aggressive financial reporting. *The Accounting Review* 84 (2): 467–496.
- Gleason, C., and L. Mills. 2002. Materiality and contingent tax liability reporting. *The Accounting Review* 77: 317–342.
- Guenther, D. 1994. Earnings management in response to corporate tax rate changes: Evidence from the 1986 tax reform act. *The Accounting Review* 69: 230–243.
- Guenther, D., and D. Young. 2000. The association between financial accounting measures and real economic activity: A multinational study. *Journal of Accounting and Economics* 29: 53–72.
- Hanlon, M. 2003. What can we infer about a firm's taxable income from its financial statements? *National Tax Journal* 56 (4): 831–863.
- . 2005. The persistence and pricing of earnings, accruals, and cash flows when firms have large book-tax differences. *The Accounting Review* 80 (1): 137–166.
- Hanlon, M., and S. Heitzman. 2010. A review of tax research. *Journal of Accounting and Economics* 50: 127–178.
- Hanlon, M., and T. Shevlin. 2005. Book-tax conformity for corporate income: An introduction to the issues. In *Tax policy and the economy*, vol. 19, 101–134. National Bureau of Economic Research.
- Hanlon, M., S. Laplante, and T. Shevlin. 2005. Evidence for the possible information loss of conforming book income and taxable income. *Journal of Law and Economics* 48: 407–442.
- Hanlon, M., E.L. Maydew, and T. Shevlin. 2008. An unintended consequence of book-tax conformity: A loss of earnings informativeness. *Journal of Accounting and Economics* 46 (2–3): 294–311.
- Hanlon, M., G.V. Krishnan, and L.F. Mills. 2012. Audit fees and book–tax differences. *Journal of the American Taxation Association* 34: 55–86.
- Hasan, I., C. Hoi, Q. Wu, and H. Zhang. 2014. Beauty is in the eye of the beholder: The effect of corporate tax avoidance on the cost of bank loans. *Journal of Financial Economics* 113: 109–130.
- Hung, M. 2001. Accounting standards and value relevance of financial statements: An international analysis. *Journal of Accounting and Economics* 30: 401–420.
- Joos, P., J. Pratt, and S. Young. 2000. *Book-tax differences and the pricing of earnings*. Working paper, Massachusetts Institute of Technology.
- Kim, J.B., O. Li, and Y. Li. 2010. *Corporate tax avoidance and bank loan contracting*. Working paper, City University of Hong Kong.
- Kuo, N.T., and C.F. Lee. 2016. A potential benefit of increasing book–tax conformity: Evidence from the reduction in audit fees. *Review of Accounting Studies* 21: 1287–1326.
- Kvaal, E., and C. Nobes. 2012. *On the definition and measurement of book-tax conformity*. Unpublished working paper, BI Norwegian Business School.
- Leuz, C., D. Nanda, and P. Wysocki. 2003. Earnings management and investor protection: An international comparison. *Journal of Financial Economics* 69: 505–527.
- Lev, B., and D. Nissim. 2004. Taxable income, future earnings, and equity values. *The Accounting Review* 79: 1039–1074.

- Lisowsky, P., L. Robinson, and A. Schmidt. 2013. Do publicly disclosed tax reserves tell us about privately disclosed tax shelter activity? *Journal of Accounting Research* 51 (3): 583–629.
- Manzon, G., and G. Plesko. 2002. The relation between financial and tax reporting measures of income. *Tax Law Review* 55: 175–214.
- Maydew, E. 1997. Tax induced earnings management by firms with net operating losses. *Journal of Accounting Research* 35: 83–96.
- McClelland, J., and L. Mills. 2007. Weighing benefits and risks of taxing book income. *Tax Notes* 114: 779–787.
- Mills, L. 1998. Book-tax differences and internal revenue service adjustments. *Journal of Accounting Research* 36 (2): 343–356.
- Mills, L., K. Newberry, and W. Trautman. 2002. *Trends in book-tax income and balance sheet differences*. Tax Notes 96, Special Report, 1109–1124.
- Moore, J. 2012. Empirical evidence on the impact of external monitoring on book-tax differences. *Advances in Accounting, Incorporating Advances in International Accounting* 28 (2): 254–269.
- Moore, J., and L. Xu. 2018. Book-tax differences and costs of private debt. *Advance in Accounting* 42: 70–82.
- Penman, S. 2001. *Financial statement analysis and equity valuation*. Boston: McGraw-Hill.
- Phillips, J., M. Pincus, and S. Rego. 2003. Earnings management: New evidence based on deferred tax expense. *The Accounting Review* 78 (2): 491–521.
- Phillips, J., M. Pincus, S. Rego, and H. Wan. 2004. Decomposing changes in deferred tax assets and liabilities to isolate earnings management activities. *Journal of the American Taxation Association* 26: 491–521. Sup.
- Plesko, G.A. 2004. Corporate tax avoidance and the properties of corporate earnings. *National Tax Journal* 57 (3): 729–737.
- Scholes, M., G. Wilson, and M. Wolfson. 1992. Firms' responses to anticipated reductions in tax rates: The Tax Reform Act of 1986. *Journal of Accounting Research* 30: 161–185.
- Shaviro, D. 2009. Internationalization of income measures and the U.S. book-tax relationship. *National Tax Journal* 62 (1): 155–167.
- Shevlin, T. 2002. Commentary: Corporate tax shelters and book-tax differences. *Tax Law Review* 55: 427–443.
- Subramanyam, K. 1996. The pricing of discretionary accruals. *Journal of Accounting and Economics* 22: 249–281.
- Tang, T. 2015. Does book-tax conformity deter opportunistic book and tax reporting? An international analysis. *European Accounting Review* 24 (3): 441–469.
- Tang, T., and M. Firth. 2011. Can book-tax differences capture earnings management and tax management? Empirical evidence from China. *International Journal of Accounting* 46: 175–204.
- Watrin, C., N. Ebert, and M. Thomsen. 2014. Book-tax conformity and earnings management: Insights from European one and two-book systems. *Journal of the American Taxation Association* 36 (2): 55–89.
- Weber, D. 2009. Do analysts and investors fully appreciate the implications of book-tax differences for future earnings? *Contemporary Accounting Research* 26 (4): 1175–1206.
- Whitaker, C. 2005. Bridging the book-tax accounting gap. *The Yale Law Journal* 115 (3): 680–726.
- Wilson, R.J. 2009. An examination of corporate tax shelter participants. *The Accounting Review* 84 (3): 969–999.
- Yin, G. 2001. Getting serious about corporate tax shelters: Taking a lesson from history. *SMU Law Review* 54: 209–237.

Evaluating Portfolio Risk Management: A New Evidence from DCC Models and Wavelet Approach

105

Rabeh Khalfaoui, Aviral Kumar Tiwari, and Xuan Vinh VO

Abstract

We analyzed the volatility dynamics of three developed markets (the UK, USA, and Japan), during the period 2003–2011, by comparing the performance of several multivariate volatility models, namely, Constant Conditional Correlation (CCC), Dynamic Conditional Correlation (DCC), and consistent DCC (cDCC) models. To evaluate the performance of models, we used four statistical loss functions on the daily Value-at-Risk (VaR) estimates of a diversified portfolio in three stock indices: FTSE 100, S&P 500, and Nikkei 225. We based on one-day ahead conditional variance forecasts. To assess the performance of the abovementioned models and to measure risks over different timescales, we proposed a wavelet-based approach, which decomposes a given time series on different time horizons. Wavelet multiresolution analysis and multivariate conditional volatility models are combined for volatility forecasting to measure the co-movement between stock market returns and to estimate daily VaR in the time-frequency space. Empirical results show that the asymmetric cDCC model of (Aielli, G. P. Consistent estimation of large scale dynamic conditional correlations, working paper n.47, University of Messina, Department of Economics,

R. Khalfaoui

Laboratoire de recherche en Économie et Gestion: LR18ES27, Faculté des Sciences Economiques et de Gestion de Sfax, University of Sfax, Sfax, Tunisia

A. K. Tiwari (✉)

Rajagiri Business School, Rajagiri Valley Campus, Kochi, India

South Ural State University, Chelyabinsk, Russian Federation

Institute of Business Research, University of Economics Ho Chi Minh City, Ho Chi Minh City, Vietnam

X. V. VO

Institute of Business Research, University of Economics Ho Chi Minh City, Ho Chi Minh City, Vietnam

e-mail: vinhvx@ueh.edu.vn

Statistics, Mathematics and Sociology, 2008) is the most preferable according to statistical loss functions under raw data. The results also suggest that wavelet-based models increase predictive performance of financial forecasting in low scales according to number of violations and failure probabilities for VaR models.

Keywords

Dynamic Conditional Correlation · Discrete Wavelet Transform · Value-at-Risk · Stock price

JEL Classification

G1 · G15 · C01 · C14

105.1 Introduction

Financial managers and traders are mostly interested in measuring market risk. The most widely used measure of risk management is the Value-at-Risk (VaR), which was introduced by Jorion (1994). Forecasting VaR is based on the volatility models. The well-known volatility model is the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986). The success of this model has subsequently led to a family of univariate and multivariate GARCH models, which can capture the different behavior in financial returns. The development of this family of models has led to the development of conditional VaR forecasts.

The literature on multivariate GARCH models is large and expanding. Engle and Kroner (1995) defined a general class of multivariate GARCH (MGARCH) models. The popular ones are diagonal VEC model of Bollerslev et al. (1988), the BEKK model of Engle and Kroner (1995). While popular, these models have limitations. See Bauwens et al. (2006) for more details. In particular, diagonal VEC lacks correlation between the variance terms, BEKK can have poorly behaved likelihood function (making estimation difficult, especially for models with more than two variables). VEC has a large number of free parameters (which make it impractical for models with more than two variables). To deal with the curse of the dimensionality, Engle (2002) proposed the dynamic conditional correlations (DCC) model, which generalizes the specification of Bollerslev (1990) by assuming a time variation of correlation matrix. Tse and Tsui (2002) defined a multivariate GARCH model, which includes time-varying correlations and yet satisfies the positive-definite condition. Ling and McAleer (2003), McAleer et al. (2009), and Caporin and McAleer (2009, 2010) proposed another family of multivariate GARCH models, which assume constant conditional correlations, and do not suffer from the problem of dimensionality, by comparison with the VEC and BEKK models. The utility of these models is that modeling conditional variances allow large shocks to one variable to affect the variances of the other variables. However, accurate prediction of the stock market volatility is still a great challenge and this phenomenon has received further attention by scholars, for instance, some recent studies such as those

of Tang et al. (2021), Akkoc and Cevcir (2019), Allard et al. (2020), Hou et al. (2019), Khalfaoui et al. (2019), and Ashfaq et al. (2020), among others. The abovementioned scholars have been employed in their research studies different volatility models, including VAR-DCC-GARCH, DCC-GARCH, DCC-MIDAS, HAR, SVAR-DCC-GARCH. Based on the HAR framework, Tang et al. (2021) conclude that oil future intraday information supports to increase the forecasting performance of the US stock market volatility. Akkoc and Cevcir (2019), highlighted a time-varying co-movement and volatility spillovers from oil and gold markets to Turkish stock market.

Another important topic in the financial econometrics is the asymmetric behavior of conditional variances. The basic idea is that negative shocks have a different impact on the conditional variance evolution than do positive ones of similar magnitude. Cappiello et al. (2006), Aielli (2008), and Palandri (2009) extended the DCC model of Engle (2002) to an asymmetric DCC model, which is a generalization of the DCC model (authors develop a model capable of allowing for conditional asymmetries not only in volatilities but also in correlations). Authors used several asymmetric versions of the DCC model. In our entry, we based on a version that allows asymmetry in conditional variance. This phenomenon was raised by Nelson (1991) in introducing the Exponential GARCH (EGARCH) model, and was also considered by Glosten et al. (1993) (GJR-GARCH models), Rabemananjara and Zakoian (1993), and Zakoian (1994) (Threshold GARCH models) for the univariate case. More recently, several empirical studies based on symmetric and asymmetric multivariate GARCH models have been employed in the area of finance. For instance, Cappiello et al. (2006) used the asymmetric generalized-DCC (AG-DCC) specification to investigate asymmetries in conditional variances and correlation dynamics of three groups of continents (Europe, Australia, and North America). They provide evidence that equity returns show strong evidence of asymmetries in conditional volatility, while little is found for bond returns. Chiang et al. (2007) applied a DCC model to nine Asian stocks and confirm a contagion effect during the Asian crisis. Ho et al. (2009) applied various multivariate GARCH models to investigate the evidence of asymmetry and time-varying conditional correlations between five sectors of Industrial Production of the USA. They provide also, strong evidence of asymmetric conditional volatility in all sectors and some support of time-varying correlations in various sectoral pairs. Buttner and Hayo (2011) used a bivariate DCC model to extract dynamic conditional correlations between European stock markets. Chang et al. (2011) employed several multivariate GARCH models (CCC, DCC, VARMA-GARCH and VARMA-AGARCH) to model conditional volatility in the returns of rubber spot and futures in major rubber futures and rubber spot Asian markets, in order to investigate volatility transmissions across these markets. Their results provide the presence of volatility spillovers and asymmetric effects of positive and negative return shocks on conditional volatility. Kenourgios et al. (2011) applied the AG-DCC model and a multivariate regime-switching Gaussian copula model to capture nonlinear correlation dynamics in four emerging equity markets (Brazil, Russia, India, and China) and two developed markets (the USA and UK). Arouri et al. (2011)) employed a VAR-GARCH

approach to investigate the return linkage and volatility transmission between oil and stock markets in Gulf Cooperation Council countries. Lahrech and Sylwester (2011) used DCC multivariate GARCH models to examine the dynamic linkage between the USA and Latin American stock markets. Their results show an increase in the degree of co-movement between Latin American equity markets and US equity ones.

In this study, we employ three multivariate GARCH models, such as the CCC model of Bollerslev (1990), the DCC model of Engle (2002), and the cDCC model of Aielli (2008). These models impose a useful structure on the many possible model parameters. However, parameters of the model can easily be estimated and the model can be evaluated and used in a straightforward way. Our empirical methodology follows a two-step approach. The first step applies these dynamic conditional correlation models to model conditional volatility in the returns of three developed stock markets (the UK, USA, and Japan stock markets), in order to examine the evidence of time-varying conditional variances and correlations between stock markets. Moreover, in order to show the asymmetric effects of positive and negative return shocks on conditional volatility the EGARCH and GJR-GARCH models of Nelson (1991) and Glosten et al. (1993), respectively, were employed for modeling univariate conditional volatility. In the second step, we reexamine the dynamic conditional correlation analysis among the three major developed stock markets through a novel approach, wavelet analysis. This technique is a very promising tool as it is possible to capture the time and frequency varying features of co-movement within a unified framework, which represents a refinement to previous approaches. This wavelet-based analysis takes into account the distinction between the short- and long-term investor. From a portfolio diversification view, there exist a kind of investors who are more interested in the co-movement of stock returns at higher frequencies (lower scales), that is, short-term fluctuations, and also, there exist a kind of investors whose focus is on the relationship at lower frequencies (higher scales), that is, long-term fluctuations. The study of the co-movement of stock market returns, i.e., dynamics of variances and correlations, across scales is crucial for risk assessment of a portfolio. In terms of risk management, a higher co-movement (higher covariances) among the assets of a given portfolio implies lower gains. According to investors or traders, evaluating the co-movement of assets is a great importance to best assess the risk of a portfolio. Several applications of wavelet studying the co-movement of stock indices have been applied, including Khalfaoui et al. (2015), Ko and Lee (2015), Sharkasi et al. (2005), Rua and Nunes (2009), Rua (2010), and Masih et al. (2010). In a more recent study, Tiwari et al. (2013, 2016), Mensi et al. (2016), Albulescu et al. (2017), Das et al. (2018), Nasreen et al. (2020), Sharif et al. (2020), Huang and Huang (2020), Choi (2020), Das and Kumar (2018), Zhang et al. (2017), Uddin et al. (2019), and Zaremba et al. (2019) have used wavelet tools to analyze the dependence between several stock markets. Sharif et al. (2020) had examined the time-varying dependence structure between oil price volatility shocks, stock market price, the novel Covid-19 pandemic, the geopolitical risk, and the economic policy uncertainty in the USA using the wavelet coherence analysis and the wavelet-based Granger causality tests. Their results uncover that the effect of the Covid-19 crisis on the geopolitical risk is

significantly higher than on the economic policy uncertainty. Further, the results document that the US stock markets initially reacted to oil price shock rather than Covid-19 events occurring from Wuhan and other regions. Huang and Huang (2020) studied the dynamics of the co-movement between China's crude oil futures and international crude oil futures (WTI and Brent) at several timescales. The authors had employed the wavelet coherence and phase relation techniques to discuss the intensity of the co-movement and the direction of the co-movement between markets as well as the lead-lag relationship of future oil prices changes. According to their methods, they found heterogeneous co-movement between China's crude oil future markets and international crude oil future markets over wavelet scales. Furthermore, the authors documented that the strength of the co-movement between future markets is weak and the direction of the co-movement is unstable across time and frequency. Choi (2020) applied the continuous wavelet analysis to revisit the impact of the economic uncertainty due to Covid-19 pandemic on the volatility of S&P 500 index. The results show that the S&P 500 volatility index was influenced by the economic policy uncertainty in terms of Covid-19 crisis.

In the same vein, Das and Kumar (2018) have extended the research of Ko and Lee (2015) by employing the multiple and partial wavelet coherence techniques to investigate the nexus between international economic policy uncertainty and the stock market prices in 11 developed economies and 6 emerging economies. Their major results revealed that the stock markets in the emerging countries were more sensitive to domestic economic policy uncertainty. In the forecasting perspective, Zhang et al. (2017) and Uddin et al. (2019) have analyzed the predictability of different time series using wavelet approaches. Overall, the results of the abovementioned scholars were crucial and shed new light upon the forecastability performance.

In this chapter, we also investigate the selection of the multivariate GARCH models used in the two approaches (with and without wavelet analysis) to identify which model has the best out-of-sample forecasting performance. The assessment of the forecast performance of these models is based on out-of-sample one-day ahead conditional matrix forecasts. However, to measure model performance we used four statistical loss functions. Hansen et al. (2003), Hansen and Lunde (2005), Becker and Clements (2008), and others, showed that evaluation of univariate volatility forecast is well understood, while for an applied point of view there are no clear guidelines available on model evaluation and selection in multivariate setting (see Laurent et al. 2011).

The empirical evidence showed that the conditional variances and correlations of the UK, USA, and Japan stock market returns were dynamic and the three markets were highly correlated. We also showed that the cDCC model of Aielli (2008) is better than the CCC and DCC models based on one-day ahead out-of-sample forecasts. With regard to wavelet-based multivariate conditional volatility approach, our findings suggested a multi-scale behavior of the three markets under study, which decompose the total spillover into three sub-spillovers and decomposes the market risk measured by VaR into wavelet VaR (WVaR) estimates. In addition, this new approach helps traders and investors to reduce risk management on their investing time horizons.

The purpose of the chapter is four-fold. First, we estimated multivariate conditional volatility for stock market returns using several recent models of multivariate conditional volatility. Second, we investigated the importance of volatility spillovers on the conditional variance across the three developed markets. Third, we focused on the forecasting performance of the multivariate conditional volatility models under study for the last 250 days of the data set. Forecast comparison is based on four different loss functions including the mean squared error, the mean absolute error, the mean absolute percentage error, and the logarithm loss error. Fourth, we proposed a wavelet-based multi-resolution analysis in order to combine between traditional multivariate conditional volatility models and wavelet decomposition. The combination of wavelet decomposition and dynamic conditional correlation models were introduced to analyze the co-movements and volatility spillovers between three developed markets on multi-scale framework, based on behavior of investors. Finally, we compared the performance of wavelet-based multivariate conditional volatility model against the traditional one for one-day ahead forecast.

The structure of the chapter is organized as follows: Sect. 105.2 presents the data used for the empirical analysis and the multivariate conditional volatility methodology. Section 105.3 reports the empirical results under raw data. Wavelet analysis is discussed in Sect. 105.4. Finally, concluding remarks are stated in Sect. 105.5.

105.2 Methodology and Empirical Specifications

105.2.1 Data

Our data on stock market prices consist of the S&P 500, FTSE 100, and NIKKEI 225 composite indices for the USA, UK, and Japan. We employed daily data over the period from 01 January 2003 to 04 February 2011. Indices are obtained from DataStream. We use daily data in order to retain a high number of observations to adequately capture the rapidity and intensity of the dynamic interactions between markets.

Returns of market i (index i) at time t are computed as $r_{i,t} = \log(P_{i,t}/P_{i,t-1}) \times 100$, where $P_{i,t}$ and $P_{i,t-1}$ are the closing prices for day t and $t-1$, respectively. The times series plots are displayed in Fig. 1.

105.2.2 Descriptive Statistics

The summary statistics of the data are given in Table 1. In panel A (Table 1), for each return series, the mean value is close to zero. For each return series the standard deviation is larger than the mean value. Each return series displays a small amount of skewness and large amount of kurtosis and the returns are not normally distributed.

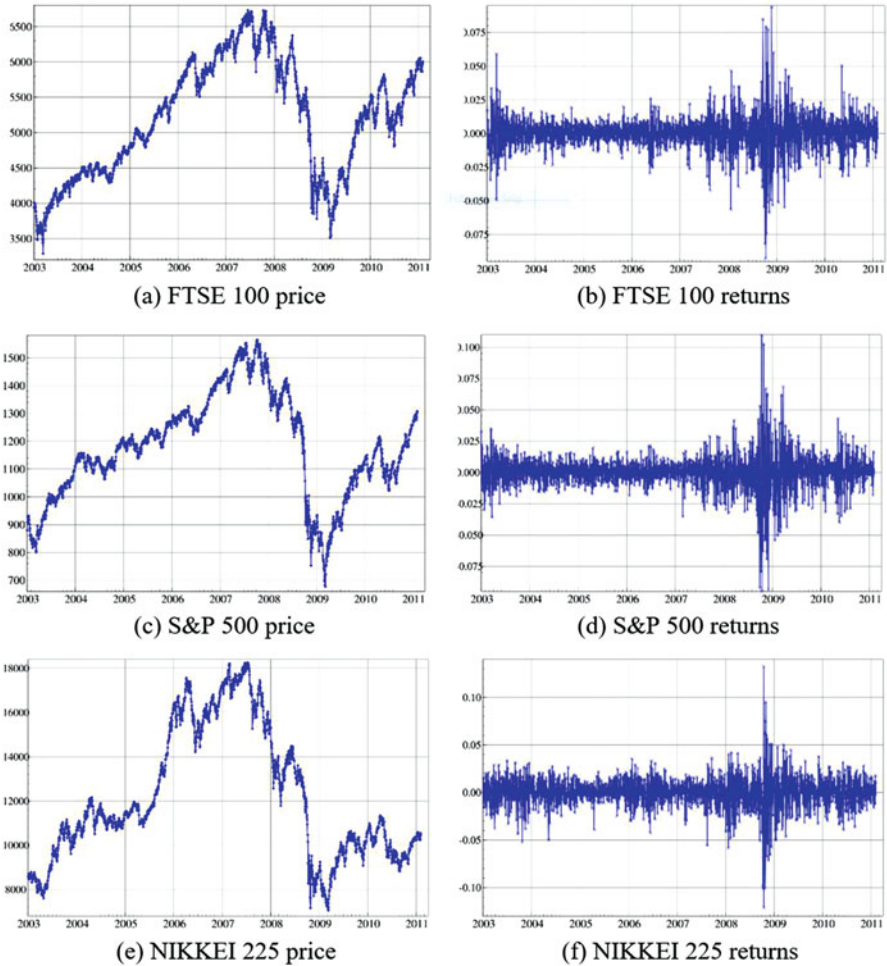


Fig. 1 Time series plots of FTSE 100, S&P 500, and NIKKEI 225 stock market indices. We plot daily level of the stock market indices (left panel) and corresponding returns (right panel) in the period covering 01 January 2003 to 04 February 2011

In panel B (Table 1), unconditional correlation coefficients in stock market index returns indicate strong pairwise correlations. The correlation between S&P 500 and FTSE 100 is positive and larger than the correlation between NIKKEI 225 and FTSE 100. This could be due to the high trade share between the two markets.

The results of the unit root tests for all sample of level prices (in logs) and returns in each market are summarized in Table 2. The Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests are used to explore the existence of unit roots in individual series. The results show that all returns are stationary.

Table 1 Descriptive statistics

Panel A: Summary statistics							
	Mean	Max	Min	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
FTSE 100	0.0199	9.384	−9.265	1.252	−0.0914	9.019	7180.19*
S&P 500	0.0187	10.957	−9.469	1.314	−0.256	11.410	11,508.39*
NIKKEI 225	0.0097	13.234	−12.111	1.544	−0.466	8.889	7048.35*
Panel B: Unconditional correlation matrix							
	FTSE 100	S&P 500	NIKKEI 225				
FTSE 100	1	0.55	0.34				
S&P 500		1	0.11				
NIKKEI 225			1				

Notes: * denotes the rejection of the null hypothesis of normality at the 1% significance level of Jarque-Bera test. The data frequency is daily and covers the period from 01 January 2003 to 04 February 2011

Table 2 Unit root tests

Prices	ADF test			Phillips-Perron test	
	None	Constant	Constant and trend	Constant	Trend
FSTE 100	0.724	−1.870	−1.920	−1.847	−1.866
S&P 500	0.644	−1.908	−1.879	−1.951	−1.912
NIKKEI 225	0.247	−1.744	−1.807	−1.670	−1.733
Panel B: Stock market returns					
Returns	ADF test			Phillips-Perron test	
	None	Constant	Constant and trend	Constant	Trend
FSTE 100	−35.171	−35.177	−35.171	−49.933	−49.924
S&P 500	−37.754	−37.758	−37.754	−52.593	−52.587
NIKKEI 225	−34.817	−34.811	−34.833	−46.359	−46.382

Notes: Entries in bold indicate that the null hypothesis is rejected at 1% significance level

105.2.3 Model Specifications

The econometric specification used in our study has two components. To model the stock market return we used a vector autoregression (VAR). To model the conditional variance, we used a multivariate GARCH model.

A VAR of order p , where the order p represents the number of lags that includes N variables can be written as the following form:

$$Y_t = \Phi_0 + \sum_{i=1}^p \Phi_i Y_{t-i} + \varepsilon_t, \quad t = 1, \dots, T \tag{1}$$

where $Y_t = (Y_{1t}, \dots, Y_{Nt})'$ is a column of observations on current values of all variables in the model, Φ_i is $N \times N$ matrix of unknown coefficients, Φ_0 is a column

vector of deterministic constant terms, $\varepsilon_t = (\varepsilon_{it}, \dots, \varepsilon_{Nt})'$ is a column vector of errors. Following Brooks (2002), the main advantage of the VAR is that there is no need to specify which variables are the endogenous variables and which are the explanatory variables because in the VAR, all selected variables are treated as endogenous variables. That is, each variable depends on the lagged values of all selected variables and helps in capturing the complex dynamic properties of the data. Note that selection of appropriate lag length is crucial. If the chosen lag length is too large relative to the sample size, the degrees of freedom will be reduced and the standard errors of estimated coefficients will be large. If the chosen lag length is too small, then the selected lags in the VAR analysis may not be able to capture the dynamic properties of the data. The chosen lag length should be free of the problem of serial correlation in the residuals. Our basic VAR will have the three stationary variables, first log differences of FTSE 100, S&P 500, and NIKKEI 225 stock market prices (will be defined in empirical section).

We focused on the modeling of multivariate time-varying volatilities. The most widely used model is DCC one of Engle (2002) which captures the dynamic of time-varying conditional correlations, contrary to the benchmark CCC model (Bollerslev 1990) which retains the conditional correlation constant. The CCC specification can be presented as:

$H_t = D_t R D_t$, where, $D_t = \text{diag} \sqrt{h_{i,t}}$ is a diagonal matrix with square root of the estimated univariate GARCH variances on the diagonal. R is the time-invariant symmetric matrix of the correlation returns with $\rho_{ii} = 1$.

In CCC model, the conditional correlation coefficients are constant, but conditional variances are allowed to vary in time.

The specification of the DCC model is as follows:

$$r_t = \mu + \sum_{s=1}^p \Phi_s Y_{t-s} + \varepsilon_t, \quad t = 1, \dots, T, \quad \varepsilon_t | \Omega_{t-1} \sim N(0, H_t) \quad (2)$$

$$\varepsilon_t = (\varepsilon_{UK,t}, \varepsilon_{US,t}, \varepsilon_{JP,t})' = H_t^{1/2} z_t, \quad z_t \sim N(0, I_3) \quad (3)$$

$$H_t = E(\varepsilon_t \varepsilon_t' | \Omega_{t-1}) \quad (4)$$

where, r_t is a 3×1 vector of the stock market index return, ε_t is the error term from the mean equations of stock market indices (Eq. 1), z_t is a 3×1 vector of *i.i.d* errors, and H_t is the conditional covariance matrix. Equation 1 can be rewritten as follows:

$$\begin{bmatrix} r_{UK,t} \\ r_{US,t} \\ r_{JP,t} \end{bmatrix} = \begin{bmatrix} \mu_{UK} \\ \mu_{US} \\ \mu_{JP} \end{bmatrix} + \begin{bmatrix} \phi_{11}^s & \phi_{12}^s & \phi_{13}^s \\ \phi_{21}^s & \phi_{22}^s & \phi_{23}^s \\ \phi_{31}^s & \phi_{32}^s & \phi_{33}^s \end{bmatrix} \begin{bmatrix} r_{UK,t-s} \\ r_{US,t-s} \\ r_{JP,t-s} \end{bmatrix} + \begin{bmatrix} \varepsilon_{UK,t} \\ \varepsilon_{US,t} \\ \varepsilon_{JP,t} \end{bmatrix}$$

To represent the Engle's (2002) DCC-GARCH model for the purpose of this study, let $r_t = (r_{UK,t}, r_{US,t}, r_{JP,t})'$ be a 3×1 vector of stock market returns, such that, r_{UK} , r_{US} , and r_{JP} are the returns of FTSE 100, S&P 500, and NIKKEI 225 indices, respectively: $r_t | \Omega_{t-1} \sim N(0, H_t)$.

The DCC-GARCH specification of the covariance matrix, H_t , can be written as:

$$H_t = D_t R_t D_t \quad (5)$$

where $D_t = \text{diag}(\sqrt{h_{UK,t}}, \sqrt{h_{US,t}}, \sqrt{h_{JP,t}})$ is 3×3 diagonal matrix of time-varying standard deviation from univariate GARCH models; i.e., $h_{i,t} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1}$, $i = UK, US, JP$, and $R_t = \{\rho_{ij}\}$ is the time-varying conditional correlation matrix.

The estimation procedure of DCC-GARCH model is based on two stages. In the first stage, a univariate GARCH model is estimated. In the second step, the vector of standardized residuals $\eta_{i,t} = r_{i,t} / \sqrt{h_{i,t}}$ is employed to develop the DCC correlation specification as follows:

$$R_t = \text{diag}(q_{11,t}^{-1/2}, \dots, q_{33,t}^{-1/2}) Q_t \text{diag}(q_{11,t}^{-1/2}, \dots, q_{33,t}^{-1/2}) \quad (6)$$

where $Q_t = (q_{ij,t})$ is a symmetric positive definite matrix. Q_t is assumed to vary according to a GARCH-type process:

$$Q_t = (1 - \theta_1 - \theta_2) \bar{Q} + \theta_1 \eta_{t-1} \eta'_{t-1} + \theta_2 Q_{t-1} \quad (7)$$

The parameters θ_1 and θ_2 are scalar parameters to capture the effects of previous shocks and previous dynamic conditional correlation on current dynamic conditional correlation. The parameters θ_1 and θ_2 are positive and $\theta_1 + \theta_2 < 1$. $\bar{Q}$ is 3×3 unconditional variance matrix of standardized residuals $\eta_{i,t}$. The correlation estimators of Eq. 7 are of the form:

$$\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}$$

In the DCC model the choice of $\bar{Q}$ is not obvious as Q_t is neither a conditional variance nor correlation. Although $E(\eta_{t-1} \eta'_{t-1})$ is inconsistent for the target, the recursion in Q_t does not have a martingale representation. See Aielli (2008) for further details. Aielli (2008) proposed the corrected Dynamic Conditional Correlation (cDCC) to evaluate the impact of both the lack of consistency and the existence of bias in the estimated parameters of the DCC model of Engle (2002). He showed that the bias depends on the persistence of the DCC dynamic parameters. Aielli (2008) showed that the lack of consistency of the three-step DCC estimator depends strictly on the persistence of the parameters driving the correlation dynamics and on the relevance of the innovations. The bias is an increasing function of both θ_1 and $\theta_1 + \theta_2$. The parameter estimates obtained from fitting DCC models are small, and close to zero for θ_1 and close to unit for $\theta_1 + \theta_2$. In order to resolve this issue, Aielli (2008) introduces the cDCC model, which has the same specification as the DCC model of Engle (2002), except for the correlation process Q_t is reformulated as follows:

$$Q_t = (1 - \theta_1 - \theta_2) \bar{Q} + \theta_1 \eta_{t-1}^* \eta'^*_{t-1} + \theta_2 Q_{t-1} \quad (8)$$

where $\eta_t^* = \text{diag}\{Q_t\}^{1/2}\eta_t$.

To investigate the asymmetric properties of stock market returns we introduce the conditional asymmetries in variance. Cappiello et al. (2006) estimate several asymmetric versions of the dynamic conditional correlation models. The version that we use is based on the following specification:

$$r_t = \mu + \sum_{s=1}^p \Phi^s r_{t-s} + \varepsilon_t, \quad t = 1, \dots, T, \quad \varepsilon_t \mid \Omega_{t-1} \sim N(0, H_t)$$

$$H_t = D_t R_t D_t, \quad \text{where } D_t = \text{diag}\left(\sqrt{h_{11,t}}, \sqrt{h_{22,t}}, \sqrt{h_{33,t}}\right)$$

$$h_{it} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \gamma_i \mathcal{I}_{\{\varepsilon_{i,t} < 0\}} \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1} \quad \text{for } i = UK, US, JP$$

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1}, \quad \text{where } Q_t^* = \text{diag}\{\sqrt{q_{11,t}}, \sqrt{q_{22,t}}, \sqrt{q_{33,t}}\}$$

$$Q_t = (1 - \theta_1 - \theta_2)\overline{Q} + \theta_1 \eta_{t-1} \eta'_{t-1} + \theta_2 Q_{t-1}$$

105.3 Empirical Results

In this section we initially employed a vector autoregressive (VAR) model to examine the relationship among stock market returns of the three developed countries. Our model is estimated on a set of stationary variables. These variables are returns in stock market prices for the UK, USA, and Japan. Table 3 reports the findings of the VAR(8) model (lags is selected using AIC criterion).

Table 3 VAR diagnostic test

	FTSE 100 [p-value]	S&P 500 [p-value]	NIKKEI 225 [p-value]
Panel A: Univariate diagnostic test			
Q(12) [p-value]	3.8854 [0.421]	5.4819 [0.241]	3.2173 [0.522]
ARCH test [p-value]	85.192 [0.000]	73.997 [0.000]	53.224 [0.000]
JB test [p-value]	811.82 [0.000]	1732.9 [0.000]	964.66 [0.000]
Std. Dev.	0.0114	0.0128	0.0123
Panel B: Multivariate diagnostic test			
Q(12) [p-value]	39.6760 [0.309]	AIC	−18.523
JB test [p-value]	2841.6 [0.000]	SIC	−18.321
Log-likelihood	19,561.61		

Notes: VAR model was estimated using eight lags (the lag was selected using AIC criterion). P-values show statistical significance of the results

105.3.1 Conditional Variance and Volatility Analysis

This subsection presents the empirical results from symmetric and asymmetric multivariate models. In the first step the univariate GARCH(1,1) model for each stock market is fitted. We model the conditional variance as a GARCH(1,1), EGARCH(1,1), and TGARCH(1,1). In the second step the symmetric multivariate GARCH(1,1) models, such as Constant Conditional Correlation (CCC), the Dynamic Conditional Correlation (DCC), the corrected Dynamic Conditional Correlation (cDCC), and the asymmetric multivariate GARCH(1,1) models, such as; aCCC, aDCC, and a-cDCC are fitted.

105.3.1.1 Symmetric and Asymmetric Univariate GARCH Analysis

Table 4 reports the model estimates (panel A) and related diagnostic tests (panel B) for the three models and for the three stock markets. Firstly, panel A of Table 4 shows that the parameters in the conditional variance equations are all statistically significant, except for the α 's in the EGARCH model for the three markets. The estimated value of β (GARCH effect) is close to unity (in all models the estimated values are greater than 0.90) and is significant at the 1% level for each model. This indicates a high degree of volatility persistence in the UK, USA, and Japan stock market returns. Secondly, results given by the GARCH(1,1) model assume that positive and negative shocks will have the same influence in conditional volatilities forecasts. In order to identify the asymmetry in conditional volatilities, we fitted the univariate EGARCH(1,1) and TGARCH(1,1) models. This asymmetry is generally referred to as a “leverage” and a “Threshold” effects. The EGARCH(1,1) model captures this “leverage” asymmetry and the TGARCH(1,1) captures this “threshold” asymmetry. The EGARCH (Exponential GARCH) model of Nelson (1991) is formulated in terms of the logarithm of the conditional variance, as in the EGARCH(1,1) model,

$$\log(h_t) = \omega + \alpha \frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} + \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta \log(h_{t-1})$$

The parametrization in terms of logarithms has the obvious advantage of avoiding nonnegativity constraints on the parameters. The TGARCH (Threshold) GARCH model or GJR-GARCH model defined by Glosten et al. (1993) and Zakoian (1994) augments the GARCH model by including an additional ARCH term conditional on the sign of the past innovation,

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \delta \mathcal{I}_{\{\varepsilon_{t-1} < 0\}} \varepsilon_{t-1}^2 + \beta h_{t-1}$$

The results showed that the asymmetric parameter γ or δ is positive and significantly different from zero at the 5% level in the EGARCH(1,1) model, indicating that UK, USA, and Japan stock markets exhibit a leverage effect with positive shocks (good news) and significantly different from zero at 1% level in the

TGARCH(1,1) model, identifying that the UK, USA, and Japan stock markets exhibit a threshold effect with positive effect. To select the adequate model for our data, we compare the three models using three criteria; such as the Log-likelihood, Akaike Information Criterion (AIC), and Schwarz Information Criterion (SIC). We showed that the asymmetric GARCH model; TGARCH(1,1) has a superior goodness of fit for the data employed. For instance, AIC in the TGARCH(1,1) is lower than in GARCH(1,1) and EGARCH(1,1) models.

Panel B of Table 4 depicts the Ljung-Box statistics computed on the standardized residuals and squared standardized residuals. We showed that all the $Q(24)$ and $Q_s(24)$ values do not reject the null hypothesis of no serial correlation in the standardized residuals and squared standardized residuals at 1% level.

105.3.1.2 Symmetric and Asymmetric Dynamic Conditional Correlation Analysis

In the second step, the symmetric and asymmetric multivariate GARCH(1,1) models were estimated in order to investigate the constant and time-varying conditional correlation in the stock markets under study. To fit these models, the standardized residuals of the univariate GARCH(1,1) models specification (discussed in the first step of our study) was employed for the estimation of the symmetric models: CCC, DCC, cDCC and asymmetric models: aCCC, aDCC, a-cDCC. The DCC and cDCC estimates of the conditional correlations between the volatilities of the FTSE 100, S&P 500, and NIKKEI 225 returns are given in Tables 7 and 8. Results showed that all coefficients are statistically significant at 1% and 5% levels.

For the CCC model (Tables 5 and 6), the correlation between FTSE 100 and S&P 500, FTSE 100 and NIKKEI 225 and S&P 500 and NIKKEI 225 are each positive and statistically significant at 1% level, and the highest correlation is between FTSE 100 and S&P 500 followed by the correlation between FTSE 100 and NIKKEI 225. This indicates the positive co-movement between the three markets. For instance, we found that the co-movements between the UK and USA are higher than the co-movements between USA and Japan. However, we remarked that estimated constant conditional correlation coefficients of the sample stock markets do not seem to be informative on dynamic linkages and co-movements between the abovementioned markets. To evaluate the dynamic (pairwise) correlation structure of the UK, USA, and Japan stock markets, we employed the DCC and cDCC models in trivariate framework.

For the DCC and cDCC models (Tables 7 and 8), the estimated parameters θ_1 and θ_2 capture the effect of lagged standardized shocks; $\eta_{t-1}\eta'_{t-1}$, and $\eta^*_{t-1}\eta^{*'}_{t-1}$ and lagged dynamic conditional correlations; Q_{t-1} , on current dynamic conditional correlations, respectively. We remarked that these parameters are significant, and this statistical significance in each market indicates the presence of time-varying stock market correlations. Following the dynamic conditional variance, the three markets under study produce similar behavior and the estimated conditional variance shows a sharp spike in the period between 2007 and 2008 (The maximum value of estimated conditional variance is in 21 October 2008 for FTSE 100 returns, in 16 October 2008

Table 4 GARCH parameter estimates for stock market indices: Raw data

	United Kingdom			United States			Japan		
	GARCH	EGARCH	TGARCH	GARCH	EGARCH	TGARCH	GARCH	EGARCH	TGARCH
Panel A: Model estimates									
ω	0.0078**	-9.2720*	0.0101*	0.0112**	-9.0095*	0.0114**	0.0218**	-8.9373*	0.0244**
α	0.0894*	1.0788	0.0080	0.0707*	2.3070	-0.0078	0.0753*	0.1645	0.0449*
β	0.9055*	0.9594*	0.9267*	0.9190*	0.9616*	0.9416*	0.9101*	0.9306*	0.9091*
γ or δ		0.1266*	0.1064*		0.0651	0.1055*		0.2179*	0.0562**
Log-likelihood	6875.8	6848.3	6900.5	6764.4	6734.5	6797.5	6497.1	6468.2	6504.5
AIC	-6.5331	-6.5050	-6.5556	-6.4272	-6.3969	-6.4577	-6.1731	-6.1437	-6.1792
SIC	-6.5250	-6.4916	-6.5449	-6.4191	-6.3834	-6.4469	-6.1650	-6.1303	-6.1685
Panel B: Diagnostic test for standardized residuals									
Q(24)	21.1221	18.6651	19.0830	39.6830	34.7351	40.3153	22.6904	18.6922	25.6935
p-value	0.6315	0.7696	0.7475	0.0231	0.0724	0.0197	0.5381	0.7681	0.3688
Q _s (24)	34.6176	37.0246	33.1795	37.9826	45.9568	31.3504	19.1083	35.0669	14.5893
p-value	0.0424	0.0235	0.0593	0.0184	0.0020	0.0891	0.6386	0.0381	0.8792

Notes: The estimates are produced by the univariate GARCH(1,1), EGARCH(1,1), and TGARCH(1,1) models. The univariate variance estimates are introduced as inputs in the estimation of CCC, DCC, and cDCC models. The estimated coefficient ω denotes the constant of the variance equation, α represents the ARCH term, and β is the GARCH coefficient. γ and δ are the asymmetric effects. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively. Q(24) and Q_s(24), respectively, represent the Ljung-Box Q statistics of order 24 computed on the standardized residuals and squared standardized residuals. Value of the estimated coefficient is multiplied by 10⁴ for the EGARCH model. Values in bold indicate the selected model

Table 5 CCC diagnostic test for standardized residuals

	Log-likelihood: -7233.3	AIC: 6.8872 SIC: 6.9194	
	FTSE 100	S&P 500	NIKKEI 225
Q(12) [p-value]	25.9838 [0.0107]	17.2378 [0.1408]	13.8918 [0.3076]
Q _s (12) [p-value]	26.1324 [0.0102]	17.1204 [0.1451]	13.3307 [0.3454]

Notes: The CCC model of Bollerslev (1990) assumes that the conditional variance for each return, h_{it} , $i = \text{UK, US, JP}$, follows a univariate GARCH process. The specification is as:

$$r_t = \Phi_0 + \sum_{s=1}^p \Phi_s r_{t-s} + \varepsilon_t, \quad \varepsilon_t \mid \Omega_{t-1} \sim N(0, H_t)$$
$$\varepsilon_t = H_t^{1/2} z_t, \quad z_t \sim N(0, \mathcal{I}_3)$$
$$H_t = E(\varepsilon_t \varepsilon_t' \mid \Omega_{t-1})$$

where $r_t = (r_{UK,t}, r_{US,t}, r_{JP,t})'$ is the vector of stock market index returns, $\varepsilon_t = (\varepsilon_{UK,t}, \varepsilon_{US,t}, \varepsilon_{JP,t})'$ is the error term from the mean equation of stock market indices (Eq. 2), z_t is 3×1 vector of i.i.d errors, and H_t is the conditional covariance matrix, which satisfies the following equation: $H_t = D_t R D_t$, where $D_t = \text{diag}(h_{UK,t}^{1/2}, h_{US,t}^{1/2}, h_{JP,t}^{1/2})$, and $R = \{\rho_{ij}\}$, for $i, j = \text{UK, US, JP}$, is the unconditional (time-invariant) correlation matrix. The off-diagonal elements of the conditional covariance matrix are given by: $H_{ij} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij}$, $i \neq j$. ***, ** and * indicate significance at 10%, 5%, and 1%, respectively. Q(24) and Q_s(24), respectively, represent the Ljung-Box Q statistics of order 24 computed on the standardized residuals and squared standardized residuals.

Table 6 CCC estimates

	FTSE 100	S&P 500	NIKKEI 225
FTSE 100	1	0.5692* (0.0149)	0.2583* (0.0212)
S&P 500		1	0.1751* (0.022)
NIKKEI 225			1

Notes: The table summarizes the estimated invariant correlations between the UK, USA, and Japan stock markets, as they are produced by CCC model. Values in parentheses are standard errors. ***, **, and * indicate significance at 10%, 5%, and 1%, respectively

for the S&P 500 returns, and in 03 November 2008 for NIKKEI 225 returns). This period is related to the US subprime financial crisis. This financial crisis led the UK, USA, and Japan capital markets to abrupt downturns, dramatically increasing systematic volatility.

We conclude that the estimates of the conditional variances based on DCC and cDCC models suggest the presence of volatility spillovers in the UK, USA, and Japan stock market returns. As shown in Fig. 2, we remarked also that the dynamic conditional correlations of the three markets under study show considerable variation, and can vary from the constant conditional correlations (ρ_{UK-US} , ρ_{UK-JP} and ρ_{US-JP}) indicating that the assumption of constant conditional correlation for all shocks to returns is not supported empirically. We stated that during the period 2003–2006, correlations between the UK and USA decreased (58% to 38%), as indicated in Fig. 2. Whereas, after 2006 we remarked a substantial increase of correlations between the UK and USA, which might be due to the Afghanistan, Iraq, and Libyan wars and the American subprime financial crisis.

Table 7 DCC estimates

	θ_1	θ_2	Log-likelihood	AIC
Panel A: Model estimates				
Coefficient	0.005218*	0.992655*	−7211.0	6.8679
Std.error	0.001416	0.002210		
t-stat (p-value)	3.685 (0.000)	449.10 (0.000)		
Panel B: Diagnostic test for standardized residuals				
	FTSE 100	S&P 500	NIKKEI	
Q(12) (p-value)	21.6339 (0.041)	16.2283 (0.181)	13.9273 (0.305)	
Qs(12 (p-value))	25.7240 (0.011)	18.5131 (0.100)	13.3440 (0.344)	

Notes: The estimates of a symmetric version of Engle (2002) DCC model are computed. The specification is as follows:

$$r_t = \Phi_0 + \sum_{s=1}^p \Phi_s r_{t-s} + \varepsilon_t, \quad \varepsilon_t \mid \Omega_{t-1} \sim N(0, H_t)$$
$$\varepsilon_t = H_t^{1/2} z_t, \quad z_t \sim N(0, \mathcal{I}_3)$$
$$H_t = D_t R_t D_t$$
$$D_t = \text{diag} \left(h_{UK,t}^{1/2}, h_{US,t}^{1/2}, h_{JP,t}^{1/2} \right)$$
$$h_{it} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1}, \quad i = UK, \quad US, \quad JP.$$

The dynamic conditional correlation R_t is a time-varying matrix defined as: $R_t = Q_t^{*-1} Q_t Q_t^{*-1}$ where $Q_t^{*-1} = \text{diag} \left(\sqrt{q_{11,t}}, \sqrt{q_{22,t}}, \sqrt{q_{33,t}} \right) \cdot Q_t = (1 - \theta_1 - \theta_2) \bar{Q} + \theta_1 \eta_{t-1} \eta_{t-1}' + \theta_2 Q_{t-1}$ where $\eta_t = D_t^{-1} \varepsilon_t \cdot \bar{Q}$ is the unconditional covariance matrix of η . The elements of H_t are $H_{ij} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij}$ where $\rho_{ij} = 1$. ***, **, and * indicate significance at 10%, 5%, and 1%, respectively. Q (24) and Qs(24), respectively, represent the Ljung-Box Q statistics of order 24 computed on the standardized residuals and squared standardized residuals

Table 8 cDCC estimates

	θ_1	θ_2	Log-likelihood	AIC
Panel A: Model estimates				
Coefficient	0.005282*	0.992716*	−7210.5	6.8674
Std.error	0.001430	0.002047		
t-stat (p-value)	3.692 (0.000)	448.80 (0.000)		
Panel B: Diagnostic test for standardized residuals				
	FTSE 100	S&P 500	NIKKEI	
Q(12) (p-value)	21.6228 (0.041)	16.3030 (0.177)	13.9019 (0.307)	
Qs(12 (p-value))	25.6085 (0.012)	18.5792 (0.099)	13.357 (0.343)	

Notes: Aielli (2008) proposed a consistent DCC model. He suggested that the DCC correlation Q_t should take a slightly different from than given in DCC model. The specification is similar to that in DCC, except for Q_t which becomes: $Q_t = (1 - \theta_1 - \theta_2) \bar{Q} + \theta_1 \eta_{t-1}^* \eta_{t-1}^{*'} + \theta_2 Q_{t-1}$, $\eta_t^* = \text{diag}(Q_t)^{1/2} \eta_t$. ***, **, and * indicate significance at 10%, 5%, and 1%, respectively. Q(24) and Qs(24), respectively, represent the Ljung-Box Q statistics of order 24 computed on the standardized residuals and squared standardized residuals

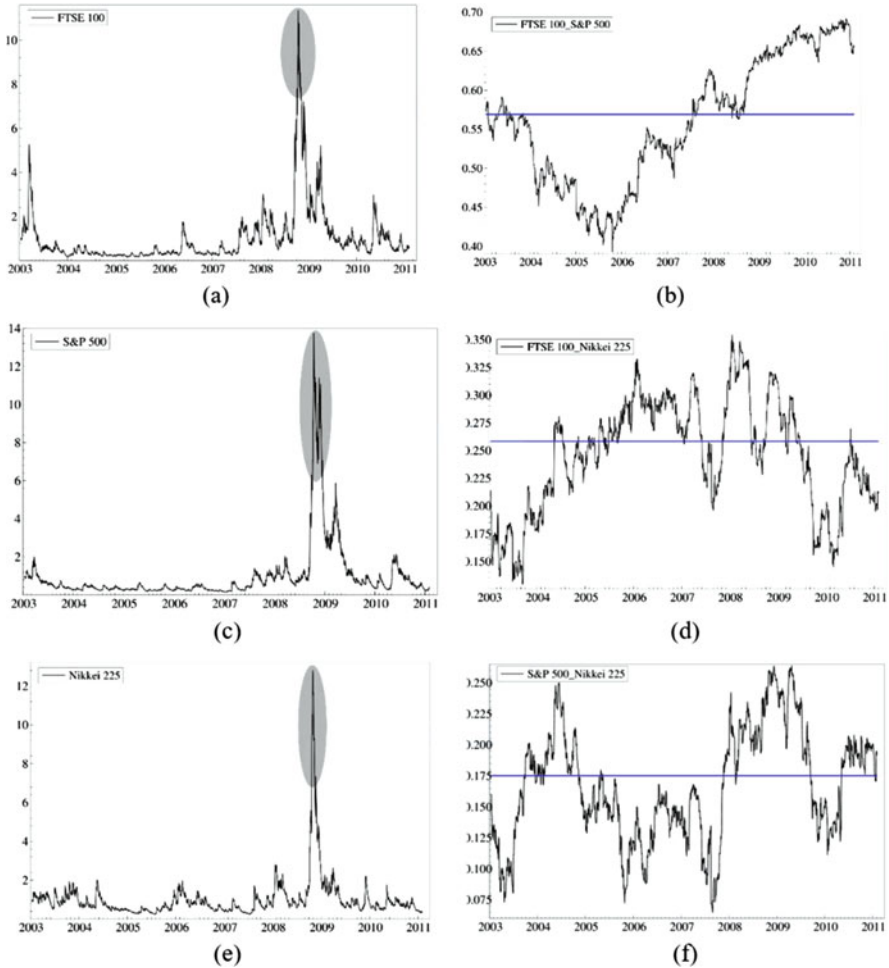


Fig. 2 Time-varying conditional variances (left column) and conditional correlations (right column) from DCC model. The pictures on the left column: (a), (c), and (e) refer to conditional variances of the UK, USA, and Japan stock market indices and those on the right column: (b), (d) and (f) report conditional correlations among the same group of stock market indices. The blue line (horizontal) is the constant conditional correlation estimate

The diagnostic tests for standardized residuals of CCC model are shown in Table 5 and those of DCC and cDCC models are shown in panel B of Tables 7 and 8. Ljung-Box $Q(12)$ and $Q_s(12)$ statistics for the residuals models indicate no serial correlation in either the standardized residuals $Q(12)$ or the squared standardized residuals $Q_s(12)$, inferring that the fitted models are appropriate for the data employed.

To investigate the asymmetry in the conditional volatility we fitted the aCCC, aDCC, and a-cDCC models. The results are reported in Table 9. In the asymmetric

Table 9 Asymmetric DCC estimates

	θ_1	θ_2	Log-likelihood	AIC	BIC
aCCC			−7169.5	6.8293	6.8696
aDCC	0.0055 (0.00162)*	0.9913 (0.00295)*	−7148.3	6.8111	6.8565
a-cDCC	0.0056 (0.00162)*	0.9913 (0.00273)*	− 7148.0	6.8108	6.8565

Notes: This table presents estimates coefficients for the asymmetric CCC [aCCC], the asymmetric DCC [aDCC] model, and the asymmetric cDCC [a-cDCC] model. The specification is similar to those in DCC and cDCC models (given in Tables 7 and 8), except of the conditional variances which follow the asymmetric TGARCH model of Glosten et al. (1993)

case, we focused on the TGARCH(1,1)-based model to model the asymmetric behavior in the conditional volatilities. The results from Table 4 indicate that TGARCH(1,1) model compared to GARCH(1,1) and EGARCH(1,1), achieves a significant improvement in the log-likelihood function, AIC and SIC. As shown in Table 4 the TGARCH(1,1) model has the lowest AIC and SIC, and maximize the log-likelihood value. For instance, the AIC values are −6.5556, −6.4577, and −6.1792 for the FTSE 100, S&P 500, and NIKKEI 225 returns, respectively. The log-likelihood values are 6900.5, 6797.5, and 6504.5 for the FTSE 100, S&P 500, and NIKKEI 225 returns, respectively. The apparent superiority of the TGARCH specification compared with EGARCH models is that the former is more robust to large shocks. This intuition is supported by Nelson and Foster (1994). The authors showed that the TGARCH model is the consistent estimator of the conditional variance of near diffusion processes. The asymmetric dynamic conditional correlation estimates are all significant at 1% level (as shown in Table 9). Based on the log-likelihood values and AIC criterion reported in the last two columns of Table 9, the asymmetric cDCC (a-cDCC) is superior to the aCCC and aDCC. Moreover, as it can be seen, the log-likelihood values and AIC criteria of aDCC and a-cDCC are nearly equivalent; hence the choice of model needs to be made on other grounds (see the following subsection).

The values in parentheses are standard errors. * indicate significance at 1% level. Log-likelihood is the log-likelihood value. AIC is the Akaike information criterion. Values in bold indicate the selected model.

105.3.2 Forecast Performance Evaluation

To evaluate the volatility forecasting performance of different multivariate GARCH models and to compare between them, further loss functions can be used. The popular statistical loss functions employed to assess the accuracy of competing models in the forecasting of volatilities over daily trading horizon are Mean squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Logarithm Loss Error (LLE). These loss metrics are expressed as follows:

$$\text{MSE}_k = \frac{1}{M} \sum_{m=1}^M (\zeta_{m,k} - \hat{\sigma}_m^2)^2 \quad (9)$$

$$\text{MAE}_k = \frac{1}{M} \sum_{m=1}^M |\zeta_{m,k} - \hat{\sigma}_m^2| \quad (10)$$

$$\text{MAPE}_k = \frac{1}{M} \sum_{m=1}^M \left| \frac{\zeta_{m,k} - \hat{\sigma}_m^2}{\hat{\sigma}_m^2} \right| \quad (11)$$

$$\text{LLE}_k = \frac{1}{M} \sum_{m=1}^M [\log(\zeta_{m,k}) - \log(\hat{\sigma}_m^2)]^2 \quad (12)$$

where M is the number of forecast data points; $\zeta_{m,k} = \kappa' H_m \kappa$ denotes the portfolio volatility forecast generated by model k for day m ; $\hat{\sigma}_m^2$ is the actual volatility on day m .

The assessment of the forecast performance of the various volatility models described in Sect. 105.2.3 is based on out-of-sample one-day ahead prediction conditional variances and correlations. This forecasting study is done by first removing the last 250 observations from our sample starts from 02 January 2003 and ends 04 February 2011. A forecast of volatilities and correlations are generated for the period $n + 1$, where n is the size of the first sample to be estimated starts 02 January 2003 and ends 19 February 2010 (the estimation is based on three models: CCC, DCC, and cDCC). The second sample, starting 03 January 2003 and ending on $n + 1$, is used to forecast the volatility and correlation of $n + 2$ based on the estimated models (CCC, DCC, and cDCC) for the second sample. The procedure of estimation and forecasting steps is repeated 250 times for the available sample from 02 January 2003 to 04 February 2011. For instance, forecasting the DCC model is as follows: the initial sample consists of the first 1862 daily observations, i.e., 02 January 2003 to 19 February 2010. The last 250 trading days constitute the sample for which we compute one-day ahead forecasts. We re-estimated the DCC model basing on the initial sample every day using a recursive method (all estimations and forecasts are computed with OxMetrics 6). After that, we computed the conditional variance and conditional correlation forecast values. These forecast values are used to compute the loss functions of the weighted portfolio defined before. More specifically, we produce the 250 one-day ahead forecasts ζ_m , where $\zeta_m, m = n + 1, \dots, n + 250$ is the forecast of the conditional variance.

The actual volatility $\hat{\sigma}_m^2$ is unobservable. The use of squared returns as a proxy for actual volatility is used in the literature (see Kang et al. 2009; Sadorsky (2006); Wei et al. 2010). In our empirical study, the proxies of actual volatilities are as follows: for forecasts based on the unfiltered return series (original stock market return series), we proxied the actual volatility by the squared returns, r_m^2 .

Table 10 Volatility forecast performance

	Symmetric models			Asymmetric models		
	CCC	DCC	cDCC	aCCC	aDCC	a-cDCC
MSE	0.16923	0.15971	0.15986	0.09237	0.08652	0.08639
MAE	0.40981	0.39822	0.39844	0.30178	0.29195	0.29170
MAPE	313.908	304.884	304.990	221.732	214.423	214.199
LLE	111.794	111.208	111.220	105.430	104.763	104.744

Notes: The table reports the loss functions over the 250 forecasting observations for the symmetric and asymmetric dynamic conditional correlation models. The MSE, MAE, MAPE, and LLE of the weighted portfolio are computed as follows: $MSE_k = M^{-1} \sum_{m=1}^M (\varsigma_{m,k} - \hat{\sigma}_m^2)^2$, $MAE_k = M^{-1} \sum_{m=1}^M |\varsigma_{m,k} - \hat{\sigma}_m^2|$, $MAPE_k = M^{-1} \sum_{m=1}^M |(\varsigma_{m,k} - \hat{\sigma}_m^2)/\hat{\sigma}_m^2|$, and $LLE_k = M^{-1} \sum_{m=1}^M [\log(\varsigma_{m,k}) - \log(\hat{\sigma}_m^2)]^2$, where M is the number of forecast data point ($M = 250$). $\varsigma_{m,k}$ denotes the volatility forecast of the portfolio constituted of three stock market indices and generated by model k for day m ; it is defined by Eq. 8. $\hat{\sigma}_m^2$ is the actual portfolio volatility on day m

Table 10 presents the forecast accuracy statistics that consist of the mean of four loss functions: MSE, MAE, MAPE, and LLE. In terms of these criteria and based on symmetric models, we find that DCC model produces smaller values than those produced by CCC and cDCC. In terms of these statistics and basing in asymmetric models, the a-cDCC model produces the smallest values. Moreover, we remarked that all loss functions values produced by asymmetric models are smaller than those calculated by symmetric ones. In summary, we can conclude that, in our forecasting study, asymmetric models produce more accurate volatility forecasts as compared to those models with symmetric conditional volatility.

105.3.3 Application to Value-at-Risk

In this section, we present a methodology allowing us to compute the VaR of diversified portfolio. A portfolio with weight κ_i in stock market index i has return as

$$r_p = \kappa' r_t \tag{13}$$

where $r_t = (r_{UK,t}, r_{US,t}, r_{JP,t})'$ denotes the vector of stock market returns of FTSE 100 ($r_{UK,t}$), S&P 500 ($r_{US,t}$), and NIKKEI 225 ($r_{JP,t}$). $\kappa' = (\kappa_{UK}, \kappa_{US}, \kappa_{JP}) = (0.27, 0.40, 0.33)$ is the weight vector of the three given indices. These weights are computed using the R function *portfolio.optim* of the *PerformanceAnalytics* R package, which computes an efficient portfolio from the given return series in the mean-variance sense. The computed portfolio has the desired expected return and no other portfolio exists, which has the same mean return, but a smaller variance.

The portfolio variances are as follows

$$\varsigma_t = \kappa' H_t \kappa \quad (14)$$

where H_t is the conditional covariance defined in Sect. 105.2.

The multivariate GARCH-based VaR estimates for q -day holding periods are computed as follows:

$$\text{VaR}_{t+l}^k(n, \alpha) = q\kappa' \mu_{t+l} + z_\alpha \sqrt{q} \varsigma_{t+l} \quad (15)$$

where the argument (q) of VaR is used to denote the time horizon, z_α denotes the corresponding quantile, which depends on the chosen distribution. For instance, when computing a 99% VaR using normal distribution, $z_\alpha = 2.33$. $\varsigma_{t+l} = (\kappa' H_{t+l} \kappa)^{1/2}$ is the square root of the daily conditional variance forecast of the portfolio generated from model k (CCC, DCC, cDCC), made at time $t + l$. μ_{t+l} is the conditional mean forecast at time $t + l$, generated from VAR(8) model.

105.3.3.1 Backtesting VaR Measures

In order to analyze our results implied by different time-varying volatilities models, we use the above estimation to compute one-day ahead out-of-sample VaR. We performed a backtesting analysis based on likelihood ratio test (Kupiec 1995; Christoffersen 1998) and dynamic quantile regression test (Engle and Manganelli 2004).

Unconditional Coverage

Kupiec (1995) developed the likelihood ratio test, LR_{uc} , as follows:

Let $n_1 = \sum_{t=1}^{n_0} \mathcal{I}_t$ be the number of days over a n_0 period that the portfolio loss was larger than the VaR estimate, where

$$\mathcal{I}_{t+l} = \begin{cases} 1 & \text{if } r_{t+l} < \text{VaR}_{t+l} \mid \Omega_t \\ 0 & \text{if } r_{t+l} \geq \text{VaR}_{t+l} \mid \Omega_t \end{cases}$$

$n_1 \sim B(n_0, \pi)$ is a binomial variable representing the number of *exceedances* or *exceptions* in the sample of length n_0 . The null hypothesis of the failure probability, π , is tested against the alternative hypothesis that the failure probability differs from π_0 . The null hypothesis for Kupiec (1995) test is $H_0 : \pi = n_1/n_0 = \pi_0$, where π is the expected proportion of exceedances, which equals the desired coverage level π_0 (usually equal to 1% and 5%). The LR_{uc} statistic is

$$LR_{uc} = -2 \ln [(1 - \pi^{n_0 - n_1}) \pi^{n_1}] + 2 \ln \left[\left(1 - \frac{n_1}{n_0}\right) \left(\frac{n_1}{n_0}\right)^{n_1} \right] \quad (16)$$

Asymptotically, $LR_{uc} \sim \chi^2(1)$.

Conditional Coverage

Christoffersen (1998) developed a likelihood statistic to test joint assumption of unconditional coverage and independence of failures. He tests the null hypothesis of independence against the alternative of first-order Markov structure of $\{\mathcal{J}_{t+1}\}$. The null hypothesis for Christoffersen (1998) test is $H_0 : \pi_{00} = \pi_{11}$, with transition matrix

$$\Pi = \begin{pmatrix} \pi_{00} & 1 - \pi_{11} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix}$$

where $\pi_{ii} = p(\mathcal{J}_{t+i} = i | \mathcal{J}_t = i)$, $i = 1, 0$. Let n_{ij} be the number of observations of $\mathcal{J}_{t+1}$ assuming value i followed by j , for $i, j = 0, 1$; and $\pi_{ij} = n_{ij} / \sum_j n_{ij}$ is the corresponding probability. Hence, we have $\hat{\pi}_{01} = n_{01} / (n_{00} + n_{01})$ and $\hat{\pi}_{11} = n_{11} / (n_{10} + n_{11})$. The statistic of the test is given by

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi^2(2) \quad (17)$$

where $LR_{ind} = -2 \ln [(1 - \pi)^{n_0 - n_1} \pi^{n_1}] + 2 \ln [(1 - \hat{\pi}_{01})^{\hat{\pi}_{01}} (1 - \hat{\pi}_{11})^{\hat{\pi}_{11}}]$
Asymptotically, $LR_{ind} \sim \chi^2(1)$.

Dynamic Quantile

Engle and Manganelli (2004) proposed the dynamic quantile (DQ) test to correct for the inefficiency in the conditional coverage test of Christoffersen (1998). They defined an indicator function $Hit_t(\alpha) = \mathcal{I}_{\{r_t < \text{VaR}_t(\alpha) | \Omega_{t-1}\}} - \alpha$ to test the VaR of long position as follows:

$$Hit_{t+1}(\alpha) = \begin{cases} 1 - \alpha & \text{if } r_{t+1} < \text{VaR}_{t+1} | \Omega_t \\ -\alpha & \text{else} \end{cases} \quad (18)$$

Engle and Manganelli (2004) suggested to test jointly the following hypothesis:

$$H_0 : \begin{cases} E(Hit_{t+1}) = 0 \\ Hit_{t+1} \text{ is uncorrelated with variables included in the information set.} \end{cases} \quad (1) \quad (2)$$

These two tests (1) and (2) can be done using artificial regression $Hit_{t+1} = X_{t+1}\beta + u_{t+1}$, where X_{t+1} is a $N \times K$ matrix whose first column is a column one and remaining columns are additional explanatory variables. footnote {We include five lags of Hit_t and the current VaR as explanatory variables. Hence, $Hit_t = (\tau, Hit_{t-1}, Hit_{t-2}, Hit_{t-3}, Hit_{t-4}, Hit_{t-5}, \text{VaR}_t(\alpha))$.

Under H_0 , Engle and Manganelli (2004) show that the dynamic quantile test statistic is given by

$$DQ = \frac{\hat{\beta}'_{OLS} X' \hat{X} \hat{\beta}_{OLS}}{\alpha(1 - \alpha)} \sim \chi^2(7) \quad (19)$$

The statistical adequacy of VaR forecasts is obtained by the previous tests. It is well-known that these tests have limited power in distinguishing among various models for VaR. Using these metrics, we cannot conclude whether an adequate model is more accurate than another one. Loss functions represent an alternative approach that can be used to compare models. Lopez (1998) suggested a loss function based on regulatory needs. He proposed measuring the accuracy of VaR forecasts on the basis of distance between observed return, r_t , and forecasted VaR $_t$ values. He defined a penalty variable as follows:

$$LF_{t+1} = \begin{cases} 1 + (r_{t+1} - \text{VaR}_{t+1}|\Omega_t)^2 & \text{if } r_{t+1} < \text{VaR}_{t+1} \mid \Omega_t \\ 0 & \text{if } r_{t+1} \geq \text{VaR}_{t+1} \mid \Omega_t \end{cases}$$

In our study we computed the loss function LF as a sum of LF_{t+1} , with $t = 0, \dots, n_0$. Table 11 depicts the out-of-sample mean daily VaR of the weighted portfolio consisting of the FTSE 100, S&P 500, and NIKKEI 225 stock market returns. We remarked that the a-cDCC model yields the lowest average daily VaR estimates at 1%, 2.5%, 5%, and 10% levels. The backtesting analysis considers a comparison over the last 250 days (from 19 February 2010 to 04 February 2011) and focuses mainly on exceptions or violations, i.e., the number of times in which the portfolio returns underperform the VaR measure, and LR_{uc} , LR_{cc} , and DQ test statistics. We also calculated the failure probability and statistic loss functions suggested by Lopez (1998). Tables 12 and 13 present unconditional, conditional, and dynamic quantile test statistics. The results showed that symmetric and asymmetric models have been rejected by the LR_{uc} , LR_{cc} , and DQ tests. Hence, these models are slow at updating the VaR values when market volatility changes rapidly.

For symmetric models, at 1% significance level, CCC, DCC, and cDCC models provide the same number of violations and failure probabilities. The same remark is shown in the case of asymmetric models. The results in Tables 12 and 13 showed that all

Table 11 Daily VaR forecasts

Model	Mean VaR α -level			
	1%	2.5%	5%	10%
Symmetric models				
CCC	1.48782	1.25373	1.05289	0.82008
DCC	1.46673	1.23595	1.03796	0.80845
cDCC	1.46716	1.23632	1.03827	0.80869
Asymmetric models				
aCCC	1.27590	1.07515	0.90292	0.70327
aDCC	1.25483	1.05739	0.88801	0.69166
a-cDCC	1.25426	1.05692	0.88761	0.69135

Notes: The table presents the out-of-sample daily VaR of the weighted portfolio consisting of three stock market indices (FTSE 100, S&P 500, and NIKKEI 225). We used quantile for normal distribution. Values in bold show the selected model for better forecasts

Table 12 Summary results for daily VaR diagnostic tests in symmetric case

	Exceedances	Fail prob. (%)	LR _{uc}	LR _{ind}	LR _{cc}	DQ	LF
1% daily VaR							
CCC	13	5.22	22.403	94.587	116.991	93.013	21.728
DCC	13	5.22	22.403	94.587	116.991	93.538	22.103
cDCC	13	5.22	22.403	94.587	116.991	93.441	22.087
2.5% daily VaR							
CCC	18	7.22	41.188	120.452	161.640		31.706
DCC	18	7.22	41.188	120.452	161.640		32.149
cDCC	18	7.22	41.188	120.452	161.640		32.131
5% daily VaR							
CCC	21	8.43	53.960	129.689	183.649	231.252	31.706
DCC	23	9.23	63.003	136.722	199.726	121.685	32.149
cDCC	23	9.23	63.003	136.722	199.726	121.586	32.131
10% daily VaR							
CCC	33	13.25	113.481	173.093	286.574	391.790	61.424
DCC	35	14.05	124.484	178.093	303.049	519.173	63.944
cDCC	35	14.05	124.484	178.093	303.049	519.110	63.924

Notes: The table presents the evaluation of out-of-sample daily VaR for the weighted portfolio (consisting of three stock market indices) generated by symmetric models. It reports test statistics over last 250 days (19 February 2010 to 03 February 2011). Fail prob.: The failure probability. LR_{uc}: The LR test of unconditional coverage. LR_{ind}: The LR test of independence. LR_{cc}: The joint test of coverage and independence. DQ: The Dynamic Quantile test. LF: The loss function. Values in bold show the selected model for better forecasts

symmetric models produce the lowest number of violations and failure probabilities. For instance, we identify 13 and 23 violations and 5.22% and 9.23% p-values for the DCC model at 1% and 5% level, respectively. While for the asymmetric DCC model, we find 18 and 31 violations and 7.22% and 12.44% p-values at 1% and 5% level, respectively.

As shown in Tables 12 and 13, we also reported the *LF* criterion, which facilitates the selection of optimal VaR model for risk manager. It is clear that CCC and aCCC models provide the lowest *LF* values at 1%, 2.5%, 5%, and 10% levels. For symmetric volatility models, the study identifies the CCC model as the best performing model for a firm, followed by cDCC and DCC models. For the asymmetric volatility models, we identify the aCCC model as the best performing model, followed by DCC and cDCC models. However, we remarked that DCC and cDCC models provide approximately the same performance.

105.4 Wavelet-Based Approach

In this section we employed a new approach called wavelet transform. The wavelet transform has two types of transform, namely, continuous wavelet transform (CWT) and discrete wavelet transform (DWT). Since most of the time series have a finite

Table 13 Summary results for daily VaR diagnostic tests in asymmetric case

	Exceedances	Fail prob. (%)	LR _{uc}	LR _{ind}	LR _{cc}	DQ	LF
1% daily VaR							
CCC	18	7.22	41.188	120.452	161.640	116.833	29.947
DCC	18	7.22	41.188	120.452	161.640	116.833	30.462
cDCC	18	7.22	41.188	120.452	161.640	116.833	30.462
2.5% daily VaR							
CCC	21	8.43	53.960	130.774	184.734	182.956	38.503
DCC	24	9.63	67.671	142.213	209.884	244.406	42.074
cDCC	24	9.63	67.671	142.213	209.884	244.437	42.083
5% daily VaR							
CCC	30	12.04	97.505	162.832	260.338	361.907	53.707
DCC	31	12.44	102.758	166.311	269.069	375.634	55.327
cDCC	31	12.44	102.758	166.311	269.069	375.646	55.338
10% daily VaR							
CCC	37	14.85	135.752	183.844	319.597	576.566	70.127
DCC	38	15.26	141.483	186.050	327.534	607.452	71.764
cDCC	18	15.26	141.483	186.050	327.534	607.488	71.776

Notes: The table presents the evaluation of out-of-sample daily VaR for the weighted portfolio (consisting of three stock market indices) generated by symmetric models. It reports test statistics over last 250 days (19 February 2010 to 03 February 2011). Fail prob.: The failure probability. LR_{uc}: The LR test of unconditional coverage. LR_{ind}: The LR test of independence. LR_{cc}: The joint test of coverage and independence. DQ: The Dynamic Quantile test. LF: The loss function. Values in bold show the selected model for better forecasts

number of observations, the discrete version of wavelet transform is used in finance and economic applications. The wavelet transform decomposes a time series in terms of some elementary functions, called, wavelets; $\psi_{u,\tau}(t) = \frac{1}{\sqrt{\tau}}\psi(\frac{t-u}{\tau})$, where $\frac{1}{\sqrt{\tau}}$ is a normalization factor, u is the translation parameter, and τ is the dilation parameter. $\psi(t)$ must fulfill several conditions (see, Gencay et al. 2002; Percival and Walden 2000): it must have zero mean, $\int_{-\infty}^{+\infty} \psi(t)dt = 0$ its square integrates to unity,

$\int_{-\infty}^{+\infty} \psi^2(t)dt = 1$, and it should also satisfy the admissibility condition, $0 < C_\psi =$

$\int_0^{+\infty} \frac{|\Psi(\lambda)|^2}{\lambda} d\lambda < +\infty$, where $\Psi(\lambda)$ is the Fourier transform of $\psi(t)$, that is $\Psi(\lambda) =$

$\int_{-\infty}^{+\infty} \psi(t)e^{-i\lambda t}dt$. Following the latter condition, we can reconstruct a time series $x(t)$.

We used a discrete wavelet transform (DWT), more specifically, we adopted the maximal overlap discrete wavelet transform (MODWT). We use the MODWT because we can align perfectly the details from the decomposition with the original time series. In comparison with the DWT, no phase shift will result in the MODWT (Gencay et al. 2002). For more information about the MODWT, please refer to Percival and Walden (2000), and Gencay et al. (2002). in a Multi-resolution Analysis

(MRA) framework. For more details, see Mallat (1989) and Percival and Walden (2000).

In our study, we sample the daily stock market returns at different scale j as follows: D_1 (2–4 days), D_2 (4–8 days), D_3 (8–16 days), D_4 (16–32 days), D_5 (32–64 days), D_6 (64–128 days), D_7 (128–256 days), D_8 (256–512 days)¹. We used Daubechies Least Asymmetric wavelet transformation of length $L = 8$ via $LA(L)$ to obtain multi-scale decomposition of the return series. The MRA yields an additive decomposition through MODWT given by

$$\begin{aligned} X(t) &= \sum_k \tilde{S}_{j,k} \phi_{j,k}(t) + \sum_j \sum_k \tilde{d}_{j,k} \psi_{j,k}(t), \quad j = 1, \dots, J \\ &= S_J(t) + D_J(t) + D_{J-1}(t) + \dots + D_1(t) \end{aligned} \quad (20)$$

where $S_J(t)$ refers to the smooth series and it represent the approximation that captures the long memory term properties, $D_j(t)$, $j = 1, \dots, J$ refers to detail series, which represent the contribution of frequency j to the original series.

The wavelet detail, $D_j(t)$, captures local fluctuations over the whole period of time series at each scale, while the wavelet smooth, $S_J(t)$, gives an approximation of the original series at scale J .

After computing the MODWT crystals (details and smooths) for every stock market return, and from the decomposed series ($D_1, \dots, D_8, S_8$) we classify the short-, medium-, and long-term series as follows: Short term = $D_1 + D_2 + D_3$; Medium term = $D_4 + D_5 + D_6$; Long term = $D_7 + D_8 + S_8$. This choice of time-horizon decomposition is used to classify three types of investors or traders, such as short-, medium-, and long-term ones, i.e., to analyze the behavior of investors among different time-horizons. Here the highest frequency component Short term, $D_1 + D_2 + D_3$, represents the short-term variations due to shocks occurring at a timescale of 2–16 days, it provides daily and weekly spillovers, the next component Medium term, $D_4 + D_5 + D_6$, represents the mid-term variations at timescale of 32 to 128 days, it defines the monthly and quarterly spillovers, and the third component Long term, $D_7 + D_8 + S_8$, represents the long-term variations of 256 days and more; it provides the annual spillovers. The main advantage of this classification is to decompose the risk and the volatility spillovers into three investment horizons. Therefore, we focus in three sub-spillovers. All market participants, such as regulators, traders, and investors, who trade in stock markets (in our study, the UK, USA, and Japan stock markets) make decisions over different timescales. In fact, due to the different decision-making timescales among investors, the time-varying volatilities and correlations of stock market indices will vary over the different timescales associated with those horizons (investment strategies).

In order to analyze the co-movements in returns and volatilities (the analysis is based on new time series: Short-, Medium-, and Long-term) between the three

¹ We decompose our time series up to scale 8 (scale $J \leq \log_2[(T-1)/(L-1)+1]$, where T is the number of observations of stock market returns ($T = 2112$), and L is the length of the wavelet filter $LA(8)$). We used the *wavelets* R package for the MODWT.

markets defined before and to investigate the dynamics and spillover effects, we applied a trivariate dynamic conditional correlation (CCC, DCC, and cDCC) model. We proceed with the same specifications presented in Sect. 105.2.3 to model the conditional mean and the conditional variances; we fitted a VAR(1)-MGARCH(1,1) to the three new series as follows:

$$\begin{cases} X_t = c_0 + AX_{t-1} + v_t \\ v_t = D_t z_t \end{cases} \quad (21)$$

where $X_t = (FTSE - Short_t, S \& P500 - Short_t, NIKKEI - Short_t)'$, c_0 is a (3×1) vector of constants, A is (3×3) coefficient matrix, v_t is (3×1) vector of error term from the mean equations of decomposed series, and z_t refers to a (3×1) vector of independently and identically distributed errors.

To ensure the stationarity of our reconstructed series (Short term, Medium term, and Long-term series), we applied the ADF and KPSS tests to our decomposed data. As shown in Table 14 all test statistics are statistically significant at 1% level, therefore indicating stationarity.

Table 15 presents estimates from two types of conditional volatility regressions: (i) a univariate GARCH model, and (ii) a univariate EGARCH model for each market and at each timescale. A GARCH(1,1) model and EGARCH(1,1) model proved adequate for capturing conditional heteroskedasticity. As shown in Table 15 the univariate GARCH models display more statistically significant coefficients than a univariate EGARCH one. In terms of GARCH model, the α and β coefficients (ARCH and GARCH effects, respectively) are positive and statistically significant in all stock markets, indicating highly persistent volatility dynamics in short-term horizon (lower scales/higher frequencies).

The timescale constant conditional correlations among the three stock market returns from the CCC model are summarized in Table 16. The results indicate that the highest constant correlation between markets is shown in medium-term horizon, which corresponds to monthly and quarterly time horizons. Therefore, the co-movements between the three markets are higher in medium horizon than co-movements in short or long horizons. For instance, the correlations between the

Table 14 Unit root tests of wavelet components

	FTSE 100		S&P 500		NIKKEI 225	
	ADF	KPSS	ADF	KPSS	ADF	KPSS
Short-term	-42.986*	0.006*	-46.451*	0.019*	-43.018*	0.008*
Medium-term	-54.386*	0.016*	-52.996*	0.013*	-49.944*	0.004*
Long-term	-33.796*	0.520*	-34.725*	0.486*	-47.833*	0.648*

Notes: The table reports the results of the augmented Dickey-Fuller (Dickey and Fuller 1979) and Kwiatkowski–Phillips–Schmidt–Shin (Kwiatkowski et al. 1992) tests. The KPSS test contains a constant and not a time trend, while the ADF test without constant and trend. The null hypothesis of ADF test is that a time series contains a unit root, $I(1)$ process, whereas the KPSS test has the null hypothesis of stationarity, $I(0)$ process. * indicate the rejection of the unit root null at 1% significance level

Table 15 Univariate GARCH(1,1) estimates for different time horizons

	ω	α	β	γ	Log-Likelihood	AIC
Short-term						
FTSE 100	0.0076**	0.1056*	0.8896*		−2638.6	2.5027
S&P 500	0.0089**	0.0757*	0.9149*		−2740.2	2.5989
NIKKEI 225	0.0188*	0.0868*	0.8992*		−3045.3	2.8881
Medium-term						
FTSE 100		0.8657*	0.1172*		574.33	−0.5419
S&P 500		0.8863*	0.0944*		658.69	−0.6268
NIKKEI 225		0.9542*	0.0014*		−156.81	0.1503
Long-term						
FTSE 100		−0.0508	1.0027*	2.5384*	2950.9	−2.7907
S&P 500		−0.0608	1.0209*	2.2234*	2953.6	−2.7932
NIKKEI 225		−0.0485	1.0232*	2.2854*	2008.5	−1.8982

Notes: The table summarizes the estimated coefficients produced by the univariate GARCH(1,1) model for short- and medium-term horizon series and EGARCH(1,1) model for long-term horizon series. The univariate variance estimates are introduced as inputs in the estimation process of the CCC, DCC, and cDCC models. The estimated coefficient ω denotes the constant term, α and β are the ARCH and GARCH terms, respectively, in the conditional variance equations, γ is the asymmetric parameter. The above estimates are for the sample period of 02 January 2003 to 04 February 2011. Significance levels at 1%, 5%, and 10% are denoted by *, **, and ***, respectively. Log-likelihood is the logarithm maximum likelihood function value. AIC is the Akaike information criterion

Table 16 CCC estimates under wavelet series

	Short-term			Medium-term			Long-term		
	FTSE 100	S&P 500	NIKKE 225	FTSE 100	S&P 500	NIKKE 225	FTSE 100	S&P 500	NIKKE 225
FTSE 100	1.000	0.527*	0.242*	1.000	0.689*	0.570*	1.000	0.650*	0.283*
S&P 500		1.000	0.130*		1.000	0.557*		1.000	0.351*
NIKKEI 225			1.000			1.000			1.000

Notes: The table reports the estimated timescale conditional correlations between the UK, USA, and Japan stock markets, as they are produced by the CCC model. * indicate significance at 1% level

UK and USA are as follows: 52.7%, 68.9%, and 65% for short, medium, and long term, respectively, and the correlations between the USA and Japan are: 13%, 55.7%, and 35.1% for short, medium, and long term, respectively. Furthermore, we remark that timescale constant correlations between the UK and USA are stronger than the others in all timescales.

Parameter estimates for the conditional variance-covariance equations in the DCC and cDCC models are reported in Table 17. Coefficients θ_1 and θ_2 reflect the ARCH and GARCH effects, respectively. The estimates parameters of the DCC and cDCC models are statistically significant at 1% level in all timescales, and the sum

Table 17 Estimates of the multivariate GARCH models (DCC and cDCC) for different time horizons

	θ_1	θ_2	Log-likelihood	AIC
DCC model				
Short-term	0.004457 (0.0017) *	0.993156 (0.0032) *	−8002.8	7.59530
Medium-term	0.719497 (0.0066) *	0.175299 (0.0078) *	14,834.1	−14.0408
Long-term	0.594683 (0.0269) *	0.400597 (0.0277) *	40,909.8	−38.7454
Cdcc model				
Short-term	0.004560 (0.0016) *	0.993199 (0.0027) *	−8002.2	7.59470
Medium-term	0.763499 (0.0055) *	0.231645 (0.0056) *	15,229.6	−14.4155
Long-term	0.831890 (0.0242) *	0.167528 (0.0244) *	43,104.5	−40.8247

Notes: This table reports the estimates parameters of DCC and cDCC models. Significance levels at 1%, 5%, and 10% are denoted by *, **, and ***, respectively. Log-likelihood is the logarithm maximum likelihood function value. AIC is the Akaike information criterion

$\theta_1 + \theta_2$ close to unity for the two models and at each scale, implying high persistent volatility in short-, medium- and long-term horizons. Based on DCC model, the degree of persistence, $\theta_1 + \theta_2$, are 0.997, 0.894, and 0.995 for short, medium, and long term, respectively. Based on cDCC model the degree of persistence is 0.997, 0.995, and 0.999 for the short, medium, and long terms, respectively. The short run persistence of shocks on dynamic conditional correlations is the greatest for high scales (0.594 and 0.831), while the largest long run persistence of shocks to conditional correlations is 0.997 (0.004 + 0.993) for low scales. We remark also that the short run persistence shocks increase by scales, while the long run persistence shocks decrease among scales.

The scale spillover effect in volatility provides strong persistence for all components (Short term, Medium term, and Long term); this phenomenon result from the trading of heterogeneous group of investors. In fact, at finest scales (Short term component), market participants are hedging strategists, speculators, and market makers. Therefore, they trade the three markets (the UK, USA, and Japan) simultaneously. Kim and In (2003) showed that speculators and market makers intensively trade to realize a quick profit (or minimize loss) over short timescales. In the intermediate scales (mid-horizon; Medium term component), the main traders are international portfolio managers who mainly follow index tracking trading strategies. Kim and In (2003) showed that trades typically occurs on a weekly to monthly basis, with little attention paid to daily prices. At high scales (Long-term component), the main traders are central banks, which operate on long-term horizons and often consider long-term economic fundamentals for their strategy.

The forecasting performance of the CCC, DCC, and cDCC models at each scale is evaluated by comparing the four statistical loss functions defined in Sect. 105.3.2. This forecasting study is based on recursive out-of-sample one-day ahead forecast of variance-covariance matrix of dynamic conditional models at each time horizon, i.e., Short-, Medium- and Long-term components. The “wavelet” volatility models are estimated based on 1863 observations corresponding to the period 02 January 2003 to 19 February 2010. The variables used in the models are: Short-, Medium-, and

Table 18 One-day out-of-sample volatility forecasts performance for different time horizons

	Forecast loss function			
	RMSE	MAE	MAPE	LLE
Panel A: Short-term				
CCC	1.488636	0.725253	5.076901	8.356644
DCC	1.488659	0.723788	5.050991	8.336478
cDCC	1.489430	0.7181159	4.958577	8.273333
Panel B: Medium-term				
CCC	0.139242	0.082996	1.001749	24.867650
DCC	0.139241	0.082996	1.001749	24.865190
cDCC	0.139241	0.082996	1.001749	24.864380
Panel C: Long-term				
CCC	0.015626	0.012123	0.999986	73.91793
DCC	0.015626	0.012123	0.999986	73.91789
cDCC	0.015626	0.012123	0.999986	73.91788

Notes: The table summarizes the four loss metrics estimates at different time horizons. The RMSE,

MAE, MAPE, and LLE are defined as: $RMSE_k(j) = \sqrt{M^{-1} \sum_{m=1}^M (\zeta_{m,k}(j) - \hat{\sigma}_m^2(j))^2}$, $MAE_k(j) = M^{-1} \sum_{m=1}^M |\zeta_{m,k}(j) - \hat{\sigma}_m^2(j)|$, $MAPE_k(j) = M^{-1} \sum_{m=1}^M |(\zeta_{m,k}(j) - \hat{\sigma}_m^2(j)) / \hat{\sigma}_m^2(j)|$ and $LLE_k(j) = M^{-1} \sum_{m=1}^M [\log(\zeta_{m,k}(j)) - \log(\hat{\sigma}_m^2(j))]^2$, where $\zeta_{m,k}(j)$ denotes the wavelet volatility forecast generated by model k for the day m and scale j and $\hat{\sigma}_m^2(j)$ is the actual portfolio volatility on day m and scale j .

Long-term components. We compare the accuracy of wavelet volatility forecasts based on:

$$\begin{aligned}
 RMSE_k(j) &= \left\{ \frac{1}{M} \sum_{m=1}^M (\zeta_{m,k}(j) - \hat{\sigma}_m^2(j))^2 \right\}^{1/2} \\
 MAE_k(j) &= \frac{1}{M} \sum_{m=1}^M |\zeta_{m,k}(j) - \hat{\sigma}_m^2(j)| \\
 MAPE_k(j) &= \frac{1}{M} \sum_{m=1}^M \left| \frac{\zeta_{m,k}(j) - \hat{\sigma}_m^2(j)}{\hat{\sigma}_m^2(j)} \right| \\
 LLE_k(j) &= \frac{1}{M} \sum_{m=1}^M [\log(\zeta_{m,k}(j)) - \log(\hat{\sigma}_m^2(j))]^2
 \end{aligned}$$

where $\zeta_{m,k}(j)$ is the wavelet volatility forecast generated by model k for day m and scale j and $\hat{\sigma}_m^2(j)$ is the actual volatility on day m at scale j . The actual wavelet volatility is not observable; therefore we define three proxies of $\hat{\sigma}_m^2(j)$, such as: (Short-Term)², (Medium-Term)², and (Long-Term)², for the short-term horizon, medium-term horizon, and long-term horizon, respectively.

Table 19 Daily WVaR forecasts

	Mean WVaR α -level			
	1%	2.5%	5%	10%
CCC model				
Short-term	1.7684	1.4902	1.2514	0.9747
Medium-term	0.0381	0.0321	0.0270	0.0210
Long-term	0.0025	0.0021	0.0018	0.0014
DCC model				
Short-term	1.7641	1.4865	1.2484	0.9723
Medium-term	0.0381	0.0321	0.0270	0.0210
Long-term	0.0025	0.0021	0.0018	0.0014
cDCC model				
Short-term	1.7455	1.4708	1.2352	0.9620
Medium-term	0.0381	0.0321	0.0270	0.0210
Long-term	0.0025	0.0021	0.0018	0.0014

Notes: The table presents the out-of-sample daily VaR of the weighted portfolio across wavelet scales. We used quantile for normal distribution

Table 18 summarizes the results of the one-day out-of-sample volatility forecast loss functions. We observe that, at low scales (Short-term component) and in terms of MAE, MAPE, and LLE criteria, the cDCC model provides better volatility forecasts, it has lower loss function values than CCC and DCC models. While, at intermediate scales (Medium-term component) and high scales (Long-term component), the three dynamic conditional correlation models show similar accuracy in one-day out-of-sample forecasts (the values of RMSE, MAE, and MAPE are equals).

Table 19 presents mean WVaR estimates for each of the dynamic conditional correlation models under 99%, 97.5%, 95%, and 90% confidence levels over the out-of-sample period from 22 February 2010, to 04 February 2011. This basic statistic can be shown as the preliminary understanding of average performance during the forecasting period before the implementation of backtesting tests and market risk loss criteria. As shown in Table 19, the WVaR estimates produced by the cDCC model at short-term horizon (low scales or high frequencies) are smaller than those of CCC and DCC models in 99%, 97.5%, 95%, and 90% confidence levels. For instance, at 99% confidence level, the average WVaR is 1.7455 for the cDCC model, while those of CCC and DCC are 1.7684 and 1.7641, respectively. We remark also that WVaR at medium-term horizon (intermediate scales) and long-term horizon (high scales or low frequencies) produced by the three models are equals, for each confidence level, 99%, 97.5%, 95%, and 90%. However, the mean WVaR decreases from lower scales to higher ones, such that, Short-term component (low scales/high frequencies) provides high WVaR estimates, followed by Medium-term (intermediate scales) component and Long-term (high scales/low frequencies) component, at all confidence levels.

To accurately compare the forecasting ability of the mentioned models in terms of WVaR,²

backtesting diagnostic tests: unconditional coverage of Kupiec (1995), conditional coverage of Christoffersen (1998), and dynamic quantile of Engle and Manganelli (2004) are introduced and used for determining the accuracy of wavelet model-based VaR measurements. We also used the regularity loss function of Lopez (1998) based on wavelet market risk (WVaR) defined before. It is defined in wavelet case as follows:

$$LF_{t+1}(j) = \begin{cases} 1 + (d_{t+1} - \text{VaR}_{t+1}(j)|\Omega_t)^2 & \text{if } d_{t+1} < \text{VaR}_{t+1}(j) | \Omega_t \\ 0 & \text{if } d_{t+1} \geq \text{VaR}_{t+1}(j) | \Omega_t \end{cases} \quad (22)$$

The statistical adequacy of the WVaR forecasts computed by the three dynamic conditional correlation models is obtained by the backtesting procedure described before: we will characterize the model as an adequate one for the volatility forecasting, if the null hypothesis cannot be rejected. The rejection of the null hypothesis indicates that the computed VaR estimates are not sufficiently accurate. Table 20 presents the summary backtesting statistics for the daily WVaR forecasts of wavelet return series. In terms of LR_{ucs} , LR_{ind} , LR_{cc} , and DQ statistics, the null hypothesis is rejected at 1%, 2.5%, and 5% significance level. According to these metrics, we cannot conclude whether an adequate multivariate model is more accurate than another one. The abovementioned backtesting tests focused on examining the accuracy of failure frequency and the independence of the failure process for VaR models. However, there are a large number of VaR models that can pass these statistical evaluation tests. How do risk managers choose among alternative VaR models? Which model will generate fewer regulatory capital requirements and induce less opportunity cost of capital? To answer these questions, we used a regulatory loss function related to market risk. Column 9 of Table 20 shows the mean loss function values. A WVaR model is preferred to another one, if it yields a lower averaging loss criterion value: $LF(j)$, defined as the sum of $LF_t(j)$, that is, $LF(j) = \sum_{t=1}^{n_0} LF_t(j)$. At 99%, 97.5%, and 95% confidence levels, $LF(j)$ statistic indicates that DCC and cDCC models provide strong WVaR performance (smaller $LF(j)$ values) than CCC model. We can note also that the three dynamic conditional correlation models have the same WVaR performance at medium-term (intermediate scales) and long-term (high scales) horizons for 99%, 97.5%, and 95% confidence levels.

²The multivariate GARCH-based WVaR estimate for one-day ahead forecasts are defined as follows:

$\text{VaR}_{t+m}^k(j) = \kappa' d_{t+m} + z_\alpha \zeta_{t+m}(j)$ where z_α denotes the normal quantile, $\kappa' d_{t+m}$ is the wavelet mean one-day ahead forecast estimate computed by forecasting VAR(1) based on wavelet components series. The wavelet weighted portfolio return is $d_t^{Pff} = \kappa' d_t$: $d_t = (d_{UK,t}, d_{US,t}, d_{JP,t})'$ denotes the vector of wavelet return components of FTSE 100 ($d_{UK,t}$), S&P 500 ($d_{US,t}$), and NIKKEI 225 ($d_{JP,t}$), $\kappa' = (\kappa_{UK}, \kappa_{US}, \kappa_{JP})$ is the vector of weights in the portfolio. $\zeta_t(j) = \kappa' H_t(j) \kappa$, where $H_t(j)$ is the wavelet conditional variance-covariance matrix defined in Sect. 105.2.3. Briefly, the wavelet conditional mean and wavelet conditional variance-covariance estimates are calculated from forecasting the model defined in Eq. 21.

Table 20 Summary results for daily WVaR diagnostic tests

	Exceedances	Fail. Prob. (%)	LR _{uc}	LR _{ind}	LR _{cc}	DQ	LF(j)
1% WVaR	CCC model						
	Short-term	7	2.81	58.791	64.325	68.020	68.020
	Medium-term	88	35.34	310.040	800.318	7404.16	96.628
	Long-term	98	39.35	329.860	901.689	9404.57	99.037
	DCC						
	Short-term	7	2.81	58.791	64.325	67.173	8.723
	Medium-term	88	35.34	310.040	800.318	7404.16	96.628
	Long-term	98	39.35	329.860	901.689	9404.57	99.037
	cDCC						
	Short-term	7	2.81	58.791	64.325	68.115	8.857
	Medium-term	88	35.34	310.040	800.318	7404.16	96.628
	Long-term	98	39.35	329.860	901.689	9404.57	99.037
2.5% WVaR	CCC model						
	Short-term	11	4.41	83.318	99.278	67.001	14.978
	Medium-term	90	3.61	312.373	818.685	7600.83	98.894
	Long-term	98	39.35	329.860	901.689	9404.57	99.044
	DCC model						
	Short-term	11	4.41	83.318	99.278	67.107	15.003
	Medium-term	90	3.61	312.373	818.685	7600.83	98.894
	Long-term	98	39.35	329.860	901.689	9404.57	99.044
	cDCC model						
	Short-term	11	4.41	83.318	99.278	67.182	15.199
	Medium-term	90	3.61	312.373	818.685	7600.83	98.894
	Long-term	98	39.35	329.860	901.689	9404.57	99.044
5% WVaR	CCC model						
	Short-term	16	6.42	110.467	143.731	101.254	23.170

(continued)

Table 20 (continued)

	Exceedances	Fail. Prob. (%)	LR _{uc}	LR _{ind}	LR _{cc}	DQ	LF(j)
Medium-term	94	37.75	538.797	314.791	853.588	7827.88	103.126
Long-term	99	39.75	580.171	330.707	910.879	9503.55	100.049
DCC model							
Short-term	16	6.42	33.264	110.467	143.731	101.254	23.204
Medium-term	94	37.75	538.797	314.791	853.588	7827.88	103.126
Long-term	99	39.75	580.171	330.707	910.879	9503.55	100.049
cDCC model							
Short-term	16	6.42	33.264	110.467	143.731	101.254	23.457
Medium-term	94	37.75	538.797	314.791	853.588	7827.88	103.126
Long-term	99	39.75	580.171	330.707	910.879	9503.55	100.049

Notes: The table presents the evaluation of out-of-sample daily WVaR for the wavelet details. It reports test statistics over the last 250 days (19 February 2010 to 03 February 2011). Exceedance: The term “exceedance” refers to an instance when portfolio losses are greater than corresponding VaR estimates in the backtest. In the literature, the terms “breaches” and “violations” are also used when referring to “exceedances.” Fail. prob.: The failure probability. LR_{uc}: The LR test of unconditional coverage. LR_{ind}: The LR test of independence. LR_{cc}: The joint test of coverage and independence. DQ: The Dynamic Quantile test. LF(j): The loss function. The respective critical values of LR_{uc} and LR_{cc} statistics at 5% significance level are 3.84 and 5.99

105.5 Conclusion

The empirical analysis in this chapter examined the co-movements and spillover effects in the stock market returns of three developed countries: the UK, USA, and Japan for the period 01 January 2003 to 04 February 2011. Three multivariate conditional correlation volatility models were used, namely, CCC model of Bollerslev (1990), DCC model of Engle (2002), and cDCC model of Aielli (2008). Empirical results show that multivariate estimates were significant for all returns in the CCC, DCC, and cDCC models. However, these models showed evidence of volatility spillovers and asymmetric effects of negative and positive shocks of equal magnitude on the conditional variances. The statistical significance of DCC and cDCC estimates indicates that the conditional correlations were dynamic. In fact, the variance-covariance analysis produced useful information on the dynamic correlations between the three developed markets, and for each pairwise series, the dynamic conditional correlations vary considerably from their respective constant correlations, implying the absence of any constant correlation between the stock markets under study. The empirical findings showed that the UK and US markets were highly correlated since the end of 2007 (the beginning of subprime crisis), followed by the UK and Japan markets and US and Japan markets. These results confirm the presence of spillover effects between pairwise stock market returns. The chapter also compared one-day ahead conditional volatility forecasts from the dynamic conditional correlation models used in the study, using 250 one-day out-of-sample forecasts, and showed that asymmetric cDCC model is preferred over the other models, according to the four used statistical loss metrics: mean squared error, mean absolute error, mean absolute percentage error, and logarithm loss error.

The chapter also combined the wavelet analysis and multivariate conditional volatility models to analyze the co-movements and volatility spillover effects in a multi-scale framework. Unlike the traditional multivariate dynamic conditional correlation volatility models, the wavelet-based dynamic conditional correlation approach allows one to decompose the spillover effect and co-movement into many sub-spillovers and sub-co-movements on various timescales according to heterogeneous groups of traders and investors. However, wavelet analysis helps investors to uncover the complex pattern of return and volatility spillovers on their own horizon and make a good hedging strategy on their risk. The findings of wavelet analysis show that multivariate estimates were significant for all wavelet time series. Moreover, wavelet-based multivariate models highlight volatility spillovers on the conditional variance for all stock markets under study. In fact, the wavelet study successfully decomposes the total spillover into sub-spillovers, respectively, for short-, mid-, and long-term horizons. The out-of-sample forecasts over wavelet scales are evaluated using four statistical loss functions and one-day ahead wavelet VaR (WVaR) forecast accuracy. The out-of-sample forecast results showed that high scales (Long-term component) provide smaller statistical loss metrics values than lower scales (Short-term component).

References

- Aielli, G. P. 2008. *Consistent estimation of large scale dynamic conditional correlations*. Working paper n. 47. University of Messina, Department of Economics, Statistics, Mathematics and Sociology.
- Akkoc, U., and I. Civcir. 2019. Dynamic linkages between strategic commodities and stock market in Turkey: Evidence from SVAR-DCC-GARCH model. *Resources Policy* 62: 231–239.
- Albulescu, C.T., D. Goyeau, and A.T. Tiwari. 2017. Co-movements and contagion between international stock index futures markets. *Empirical Economics* 52 (4): 1529–1568.
- Allard, A.-F., L. Iania, and K. Smedts. 2020. Stock-bond return correlations: Moving away from “one- frequency-fits-all” by extending the DCC-MIDAS approach. *International Review of Financial Analysis* 71: 101557.
- Arouri, M.E.H., A. Lahiani, and D.K. Nguyen. 2011. Return and volatility transmission between world oil prices and stock markets of the gcc countries. *Economic Modelling* 28 (4): 1815–1825.
- Ashfaq, S., Y. Tang, and R. Maqbool. 2020. Dynamics of spillover network among oil and leading Asian oil trading countries’ stock markets. *Energy* 207: 118077.
- Bauwens, L., S. Laurent, and J.V.K. Rombouts. 2006. Multivariate garch models: A survey. *Journal of Applied Econometrics* 21 (1): 79–109.
- Becker, R., and A. Clements. 2008. Are combination forecasts of s&p 500 volatility statistically superior? *International Journal of Forecasting* 24 (1): 122–133.
- Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 (3): 307–327.
- . 1990. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *The Review of Economics and Statistics* 72 (3): 498–505.
- Bollerslev, T., R.F. Engle, and J.M. Wooldridge. 1988. A capital asset pricing model with time varying covariances. *The Journal of Political Economy* 96 (01): 116–131.
- Brooks, C. 2002. *Introductory econometrics for finance*. Cambridge University Press.
- Buttner, D., and B. Hayo. 2011. Determinants of european stock market integration. *Economic Systems* 35: 574–585.
- Caporin, M., and M. McAleer. 2009. Do we really need both bekk and dcc? A tale of two covariance models, Documentos del Instituto Complutense de Analisis Economico 0904, Universidad Complutense de Madrid, Facultad de Ciencias Economicas y Empresariales.
- . 2010. *Do we really need both bekk and dcc? A tale of two multivariate garch models*. Working papers in economics 10/06, University of Canterbury, Department of Economics and Finance.
- Cappiello, L., R.F. Engle, and K. Sheppard. 2006. Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics* 04 (04): 537–572.
- Chang, C.-L., T. Khamkaew, M. McAleer, and R. Tansuchat. 2011. Modelling conditional correlations in the volatility of asian rubber spot and futures returns. *Mathematics and Computers in Simulation* 81 (7): 1482–1490.
- Chiang, T.C., B.N. Jeon, and H. Li. 2007. Dynamic correlation analysis of financial contagion: Evidence from asian markets. *Journal of International Money and Finance* 26 (7): 1206–1228.
- Choi, S.Y. 2020. Industry volatility and economic uncertainty due to the COVID-19 pandemic: Evidence from wavelet coherence analysis. *Finance Research Letters* 37: 101783.
- Christoffersen, P.F. 1998. Evaluating interval forecasts. *International Economic Review* 39 (4): 841–862.
- Das, D., and S.B. Kumar. 2018. International economic policy uncertainty and stock prices revisited: Multiple and partial wavelet approach. *Economic Letters* 164: 100–108.
- Das, D., M. Kannadhasan, A.K. Tiwari, and K.H. Al-Yahyaee. 2018. Has co-movement dynamics in emerging stock markets changed after global financial crisis? New evidence from wavelet analysis. *Applied Economics Letters* 25 (20): 1447–1453.
- Dickey, D.A., and W.A. Fuller. 1979. Distribution of the estimators for autoregressive time series with unit root. *Journal of the American Statistical Association* 74: 427–431.

- Engle, R.F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (4): 987–1007.
- . 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20: 339–350.
- Engle, R.F., and K.F. Kroner. 1995. Multivariate simultaneous generalized arch. *Econometric Theory* 11 (01): 122–150.
- Engle, R.F., and S. Manganelli. 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22: 367–381.
- Gencay, R., F. Selcuk, and B. Whitcher. 2002. In: *An introduction to wavelets and other filtering methods in finance and economics*. Academic-Press.
- Glosten, L.R., R. Jagannathan, and D.E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Technical Report* 157.
- Hansen, P.R., and A. Lunde. 2005. A forecast comparison of volatility models: Does anything beat a garch(1,1)? *Journal of Applied Econometrics* 20 (7): 873–889.
- Hansen, P. R., A. Lunde, and J. M. Nason. 2003. *Choosing the best volatility models: The model confidence set approach*. Working paper 2003-28, Federal Reserve Bank of Atlanta.
- Ho, K.-Y., A.K. Tsui, and Z. Zhang. 2009. Volatility dynamics of the us business cycle: A multivariate asymmetric garch approach. *Mathematics and Computers in Simulation* 79: 2856–2868.
- Hou, Y., S. Li, and F. Wen. 2019. Time-varying volatility spillover between Chinese fuel oil and stock index futures markets based on a DCC-GARCH model with a semi-nonparametric approach. *Energy Economics* 83: 119–143.
- Huang, X., and S. Huang. 2020. Identifying the comovement of price between China's and international crude oil futures: A time-frequency perspective. *International Review of Financial Analysis* 72: 101562.
- Jorion, P. 1994. Risk and turnover in the foreign exchange market. *University of Chicago Press* NBER: 19–40.
- Kang, S.H., S.-M. Kang, and S.-M. Yoon. 2009. Forecasting volatility of crude oil markets. *Energy Economics* 31 (1): 119–125.
- Kenourgios, D., A. Samitas, and N. Paltalidis. 2011. Financial crises and stock market contagion in a multivariate time-varying asymmetric framework. *Journal of International Financial Markets, Institutions and Money* 21 (1): 92–106.
- Khalfaoui, R., and M. Boutahar. 2012. Portfolio risk evaluation: An approach based on dynamic conditional correlations models and wavelet multi-resolution analysis. fflhshs-00793068. AMSE Working Papers 1208.
- Khalfaoui, R., M. Boutahar, and H. Boubaker. 2015. Analyzing volatility spillovers and hedging between oil and stock markets: Evidence from wavelet analysis. *Energy Economics* 49: 540–549.
- Khalfaoui, R., S. Sarwar, and A.K. Tiwari. 2019. Analysing volatility spillover between the oil market and the stock market in oil-importing and oil-exporting countries: Implications on portfolio management. *Resources Policy* 62: 22–32.
- Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54 (1–3): 159–178.
- Kim, S., and F.H. In. 2003. The relationship between financial variables and real economic activity: Evidence from spectral and wavelet analyses. *Studies in Nonlinear Dynamics & Econometrics* 7 (4): 1–18.
- Ko, J.-H., and C.-M. Lee. 2015. International economic policy uncertainty and stock prices: Wavelet approach. *Economic Letters* 134: 118–122.
- Kupiec, P.H. 1995. Techniques for verifying the accuracy of risk measurement models. Finance and Economics Discussion Series 95–24, Board of Governors of the Federal Reserve System (U.S.).

- Lahrech, A., and K. Sylwester. 2011. U.S. and Latin American stock market linkages. *Journal of International Money and Finance* 30: 1341–1357.
- Laurent, S., J.V.K. Rombouts, and F. Violante. 2011. On the forecasting accuracy of multivariate garch models. *Journal of Applied Econometrics* 27 (6): 934–955.
- Ling, S., and M. McAleer. 2003. Asymptotic theory for a vector Arma-Garch model. *Econometric Theory* 19 (02): 280–310.
- Lopez, J.A. 1998. Methods for evaluating value-at-risk estimates. *Economic Policy Review* 9802: 119–124.
- Mallat, S.G. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11: 674–693.
- Masih, M., M. Alzahrani, and O. Al-Titi. 2010. Systematic risk and time scales: New evidence from an application of wavelet approach to the emerging gulf stock markets. *International Review of Financial Analysis* 19 (1): 10–18.
- McAleer, M., S. Hoti, and F. Chan. 2009. Structure and asymptotic theory for multivariate asymmetric conditional volatility. *Econometric Reviews* 28 (5): 422–440.
- Mensi, W., S. Hammoudeh, and A.K. Tiwari. 2016. New evidence on hedges and safe havens for Gulf stock markets using the wavelet-based quantile. *Emerging Markets Review* 28 (C): 155–183.
- Nasreen, S., S.A.A. Naqvi, A.K. Tiwari, S. Hammoudeh, and S.A.R. Shah. 2020. A wavelet-based analysis of the co-movement between sukuk bonds and shariah stock indices in the GCC region: Implications for risk diversification. *Journal of Risk and Financial Management* 13 (4): 1–21.
- Nelson, D.B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59 (2): 347–370.
- Nelson, D.B., and D.P. Foster. 1994. Asymptotic filtering theory for univariate arch models. *Econometrica* 62 (1): 1–41.
- Palandri, A. 2009. Sequential conditional correlations: Inference and evaluation. *Journal of Econometrics* 153 (2): 122–132.
- Percival, D.B., and A.T. Walden. 2000. *Wavelet methods for time series analysis*. Cambridge University Press.
- Rabemananjara, R., and J.M. Zakoian. 1993. Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics* 8 (1): 31–49.
- Rua, A. 2010. Measuring comovement in the time-frequency space. *Journal of Macroeconomics* 32 (2): 685–691.
- Rua, A., and L.C. Nunes. 2009. International comovement of stock market returns: A wavelet analysis. *Journal of Empirical Finance* 16 (4): 632–639.
- Sadorsky, P. 2006. Modeling and forecasting petroleum futures volatility. *Energy Economics* 28 (4): 467–488.
- Sharif, A., C. Aloui, and L. Larisa Yarovaya. 2020. COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: Fresh evidence from the wavelet-based approach. *International Review of Financial Analysis* 70: 101496.
- Sharkasi, A., H.J. Ruskin, and M. Crane. 2005. Interrelationships among international stock market indices: Europe, asia and the americas. *International Journal of Theoretical and Applied Finance* 8 (5): 603.
- Tang, Y., X. Xiao, M.I.M. Wahab, and F. Ma. 2021. The role of oil futures intraday information on predicting US stock market volatility. *Journal of Management Science and Engineering* 6 (1): 64–74.
- Tiwari, A.K., A. Billah Dar, N. Bhanja, and A. Shah. 2013. Stock market integration in Asian countries: Evidence from wavelet multiple correlations. *Journal of Economic Integration* 28: 441–456.
- Tiwari, A.K., M.I. Mutascu, and C.T. Albulescu. 2016. Continuous wavelet transform and rolling correlation of European stock markets. *International Review of Economics & Finance* 42 (C): 237–256.

- Tse, Y.K., and A.K.C. Tsui. 2002. A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics* 20 (3): 351–362.
- Uddin, G.S., R. Gençay, S. Bekiros, and M. Sahamkhadam. 2019. Enhancing the predictability of crude oil markets with hybrid wavelet approaches. *Economic Letters* 182: 50–54.
- Wei, Y., Y. Wang, and D. Huang. 2010. Forecasting crude oil market volatility: Further evidence using garch-class models. *Energy Economics* 32 (6): 1477–1484.
- Zakoian, J.-M. 1994. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18 (5): 931–955.
- Zaremba, A., Z. Umar, and M. Mikutowski. 2019. Inflation hedging with commodities: A wavelet analysis of seven centuries worth of data. *Economics Letters* 181: 90–94.
- Zhang, K., R. Gençay, and M.E. Yazgan. 2017. Application of wavelet decomposition in time-series forecasting. *Economic Letters* 158: 41–46.



Cash Conversion Cycle and Corporate Performance: Global Evidence

106

Chong-Chuo Chang

Abstract

Previous studies have seldom explored issues regarding liquidity management; hence, we conduct a global empirical analysis of the relationship between the cash conversion cycle (CCC) and corporate performance by adopting enterprises from different countries as samples. We observe a negative relationship between the CCC and firm's profitability and value, supporting that an aggressive working capital policy can enhance corporate performance; however, this effect reduces or reverses when firms exist at the lower CCC level. Results remain identical after considering endogenous problems, changes in macroeconomic environments, economic development status, financial crises, corporate governance, and financial constraints.

Keywords

Global market · Cash conversion cycle · Liquidity management · Corporate performance

JEL Classification

G15 · G30 · G32 · G33 · G34

Note: I sincerely thank the permission of Elsevier. This chapter was published in International Review of Economics and Finance, Vol. 56, Chong-Chuo Chang, Cash Conversion Cycle and Corporate Performance: Global Evidence, 568-581, Copyright Elsevier (2018).

C.-C. Chang (✉)

Department of Banking and Finance, National Chi Nan University, Puli Township, Taiwan

e-mail: aaron@ncnu.edu.tw

© Springer Nature Switzerland AG 2022

C.-F. Lee, A. C. Lee (eds.), *Encyclopedia of Finance*,
https://doi.org/10.1007/978-3-030-91231-4_109

2597

106.1 Introduction

Finance theory discussion is generally related to one of the following categories: capital budgeting, capital structure, dividend policy, or working capital management (WCM). Although WCM is vital because of its impact on a firm's profitability and risk, and consequently its value (Smith 1980), it has received less attention than the other aforementioned categories. Jose et al. (1996) indicate that the day-to-day management of a firm's short-term assets and liabilities plays a crucial role in its success. Therefore, although WCM is short-term financial management, it often becomes a genuine source of profit. Kroes and Manikas (2014) suggest that cash flow management is a critical element of a firm's operational strategies. Working capital investment involves a trade-off between profitability and risk, and the balance between both factors is essential. Firms may have an optimal level of working capital that maximizes their value (Deloof 2003; Howorth and Westhead 2003). Decisions that can increase profitability can also increase risk; conversely, decisions that focus on risk reduction may reduce potential profitability (Filbeck and Krueger 2005; García-Teruel and Martínez-Solano 2007; Boisjoly et al. 2020; Sensini and Vazquez 2021). Tong and Wei (2011), He and Xiong (2012), Wang (2019), and Jalal and Khaksari (2020) also highlight the importance of WCM. De Vito and Gómez (2020) find that specifically during the COVID-19 pandemic, average firms with partial operating flexibility would exhaust their cash holdings in approximately two years when the environment is highly unfavorable. Therefore, during such a financial crisis, using firms' assets and liabilities efficiently is one of the appropriate solutions to mitigate financial problems (Zimon and Tarighi 2021).

Related literature suggests that an aggressive WCM policy can enhance a firm's performance. If the accounts receivable collection period is too long, the firm may face the risk of liquidity and payment recovery. Similarly, the firm may lose its inventory-carrying cost if the inventory conversion period is excessively increased. Increasing the payable deferral period may result in reduced payment stress. In addition, maintaining a high level of working capital leads to an opportunity cost if the firm relinquishes more profitable investments. Therefore, several studies have indicated that a reduced cash conversion cycle (CCC) can improve operating performance. For example, Hager (1976), Kamath (1989), Jose et al. (1996), Shin and Soenen (1998), Wang (2002), Deloof (2003), García-Teruel and Martínez-Solano (2007), Raheman and Nasr (2007), Uyar (2009), Baños-Caballero et al. (2012), Lee (2015), Oseifuah and Gyekye (2016), and Lin and Lin (2021) all indicate that an aggressive liquidity policy can enhance a firm's profitability and value. Furthermore, Soenen (1993) documented that a long CCC might be a primary reason for bankruptcy.

Other related studies have suggested a different viewpoint; that is, a firm's performance can be improved by a conservative WCM policy. Baños-Caballero et al. (2010), and Manikas and Patel (2016) asserted that a longer CCC may increase a firm's sales and profitability for several reasons: first, a firm can increase its sales by extending a higher trade credit that helps the firm to strengthen its relationships

with its customers (Ng et al. 1999). Second, larger inventories can prevent interruptions in the production process and loss of business because of the scarcity of products. In terms of accounts payables, companies may take advantage of crucial discounts for early payments if they reduce supplier financing (Ng et al. 1999; Wilner 2000). According to Czyzewski and Hicks (1992), firms with abundant cash can produce higher than average returns on assets. Afza and Nazir (2008) observe a negative relationship between a firm's profitability measures and the aggressiveness of its working capital investment; a firm yields negative returns if an aggressive working capital policy is adopted.

Based on the aforementioned findings, empirical studies on liquidity management have yielded mixed results. We conclude that the reason for this mixed result is that these studies have not conducted sufficiently thorough examinations and have not considered changes in macroeconomic environments, economic development status, financial crises, corporate governance, financial constraints, and endogeneity problems. Smith (1987), Blinder and Maccini (1991), Carpenter et al. (1994), Kashyap et al. (1994), and Michaelas et al. (1999) all indicate that changes in macroeconomic environments influence corporate liquidity. Klapper (2006) observes that the economic development status influences a business' accounts receivable by changing the credit policy. According to Céspedes et al. (2010), undeveloped financial markets and economies are volatile and allow few financing options for firms, which may influence decisions related to WCM. Campello et al. (2010) suggest that financial crises affect financial constraints and unconstrained corporate liquidity management.

The divergence in corporate governance, financial constraints, and endogeneity problems may also influence the relationship between liquidity and firm performance. Hail and Leuz (2006) observe that firms in countries with strong legal protection for investors tend to enjoy lower equity costs than firms in countries with weak legal protection for investors do. Chen et al. (2009) document that firms with strong firm-level corporate governance have lower capital costs, particularly those in countries with a weak legal protection. Shleifer and Wolfenzon (2002), Almeida et al. (2011), and Kusnadi and Wei (2011) all indicate that corporate governance influences capital costs and the changes in a firm's cash management policy. Riddiough and Wu (2009) identify substantial differences between the investment and liquidity management policies of firms and found that more (less) financially constrained firms exhibit high (low) investment and liquidity management sensitivity to variables that are measures of financial market friction. Ang and Smedema (2011) observe that firms do not always prepare for future recession because of financial constraints and low quantities of cash. According to Petersen and Rajan (1997), Shin and Soenen (1998), Opler et al. (1999), Wang (2002), Chiou et al. (2006), Bates et al. (2009), and Baños-Caballero et al. (2010), a firm's profitability and value also influence the WCM. The relationship between the CCC and corporate performance may suffer from endogeneity problems.

In the present study, we conducted a global empirical analysis of enterprises from different countries to investigate the relationship between WCM and firm's performance. We adopted the CCC as a proxy for WCM. To obtain robust results, we

considered endogenous problems, changes in macroeconomic environments, economic development status, financial crises, corporate governance, and financial constraints. The empirical results indicate that the CCC exhibits a negative relationship with firm's profitability and value, supporting that an aggressive working capital policy can enhance corporate performance; however, this effect reduces or reverses when firms exist at the lower CCC level. The results hold after accounting for various robustness checks.

The remainder of this chapter is arranged as follows: Sect. 106.2 describes the data and methodology; Sect. 106.3 presents the main results; and Sects. 106.4 and 106.5 consider endogeneity and robustness checks. The findings are summarized in Sect. 106.6.

106.2 Data and Methodology

106.2.1 Data

In this study, we conducted a global empirical analysis of the relationship between the CCC and corporate performance by adopting enterprises from different countries as samples. We apply financial statements and the market value of sample enterprises obtained from the Compustat Global Vantage database for the period of 1994–2011. Macroeconomic data are obtained from the World Bank database. We exclude firms with any segment in the financial industry (SIC 6000–6999) or the utility industry (SIC 4900–4999). To mitigate the effects of outliers and errors in the data, we omit the top and bottom one percentiles of all regression variables and firms with negative total assets, liabilities, and operating revenue account balances. The final sample includes 46 countries, 31,612 companies, and 266,547 firm-year observations.

106.2.2 Methodology

Following Soenen (1993), Deloof (2003), Padachi (2006), García-Teruel and Martínez-Solano (2007), and Baños-Caballero et al. (2010), we adopt the CCC as a proxy for WCM and a pooled ordinary least squares regression model to investigate the relationship between the CCC and corporate performance by adopting enterprises from different countries as samples. The specifications of the model are as follows:

$$\begin{aligned} IndAdjROA_{i,t} = & \beta_0 + \beta_1 IndAdjCCC_{i,t} + \beta_2 IndAdjCCC_{i,t} \times LowCCC_{i,t} \\ & + \beta_3 SIZE_{i,t} + \beta_4 DIV_{i,t} + \beta_5 CAPEXP_{i,t} + \beta_6 LEV_{i,t} + \beta_7 LagROA_{i,t} \\ & + \beta_8 RDR_{i,t} + \beta_9 STDROA_{i,t} + \beta_{10} MB_{i,t} + Industry\ dummies \\ & + Country\ dummies + Year\ dummies + \varepsilon_{i,t} \end{aligned}$$

(1)

$$\begin{aligned}
IndAdjTobin's\ Q_{i,t} = & \beta_0 + \beta_1 IndAdjCCC_{i,t} + \beta_2 IndAdjCCC_{i,t} \times LowCCC_{i,t} \\
& + \beta_3 SIZE_{i,t} + \beta_4 DIV_{i,t} + \beta_5 CAPEXP_{i,t} + \beta_6 LEV_{i,t} \\
& + \beta_7 LagROA_{i,t} + \beta_8 RDR_{i,t} + \beta_9 STDROA_{i,t} \\
& + Industry\ dummies + Country\ dummies \\
& + Year\ dummies + \varepsilon_{i,t}
\end{aligned} \tag{2}$$

where i denotes the firm, and t denotes the year. The CCC is calculated by adding the number of days of accounts receivable to the number of days of inventory and subtracting the number of days of accounts payable. The number of days of accounts receivable is calculated as the average accounts receivable divided by revenue per day. The number of days of inventory is calculated as the average inventory divided by the cost of goods sold per day. The number of days of accounts payable is calculated as the average accounts payable divided by the cost of goods sold per day. A shorter (longer) CCC indicates less (more) time between the outlay of cash and cash recovery, indicating that a firm is more likely to adopt an aggressive (conservative) WCM policy. The industry-adjusted CCC (*IndAdjCCC*) is calculated by subtracting the CCC from the industry median CCC in the corresponding year. We employ Fama-French 49-industry classification to group firms into industries. Following Aktas et al. (2015), we added an interaction term between the industry-adjusted CCC dummy and the industry-adjusted CCC ($IndAdjCCC \times LowCCC$) in the model. The industry-adjusted CCC dummy variable (*LowCCC*) equals 1 if the industry-adjusted CCC is negative and 0 otherwise. Return on assets (ROA) is a variable calculated by dividing the net income by the total assets. Tobin's Q is the ratio of the market value of equity added to the book value of debt, divided by the book value of the total assets.¹ *IndAdjROA* and *IndAdjTobin's Q* are the industry-adjusted ROA and Tobin's Q, respectively. *IndAdjROA* (*IndAdjTobin's Q*) is the ROA (Tobin's Q) subtracted from the industry median ROA (Tobin's Q) in the corresponding year.²

In accordance with related literature, we consider a set of control variables. Firm size (*Size*) is defined as the natural logarithm of the market value of equity (Core et al. 2006). Payout ratio (*DIV*) is defined as the ratio of dividends divided by the operating revenues (Lie 2005). *CAPEXP* denotes the ratio of capital expenditure and other investments divided by the total assets (McConnell and Muscarella 1985). Leverage (*LEV*) is defined as the ratio of the total debt divided by the total assets (Cho 1998; González 2013; Lin and Fu 2017; Pombo and Taborda 2017). *LagROA*

¹However, Peters and Taylor (2017) indicate that the OLS regressions exist two problems. The slopes on Tobin's Q are biased due to measurement error in Tobin's Q. Moreover, the OLS R^2 depends not just on how well Tobin's Q explains investment, but also on how well Tobin's Q proxies explain the true, unobservable Tobin's Q. To deal with these problems, Peters and Taylor (2017) reestimate the investment models using the Erickson et al. (2014) cumulant estimator, and their new Tobin's Q proxy is closer to the true Tobin's Q.

²We also perform regressions for dependent variable ROA and Tobin's Q (independent variable CCC), and obtain similar results. To save space, the results of dependent variable ROA and Tobin's Q (independent variable CCC) are not tabulated.

denotes the ROA of the previous year (Kim 2005; Lskavyan and Spatareanu 2006). *RDR* is the ratio of research and development expenditure divided by the total assets (Agrawal and Knoeber 1996; Morck et al. 1988; Doong et al. 2011). *STDROA* is the standard deviation of the ROA over the preceding 5-year period (Core et al. 1999). *MB* denotes the ratio of the market value of equity divided by the book value of equity (Core et al. 2006). We also account for time-invariant industry heterogeneity, time-invariant country heterogeneity, and time trends with a vector of industry-fixed effects, country-fixed effects, and year dummies (*Industry dummies*, *Country dummies*, and *Year dummies*, respectively). *Industry dummies*, *Country dummies*, and *Year dummies* denote the different industries, countries, and years presented in our sample, respectively. We also adjusted the standard errors for heteroskedasticity and autocorrelation using Newey and West (1987) correction.

106.3 Empirical Results

106.3.1 Preliminary Findings

106.3.1.1 Sample Description

Our sample includes 46 countries, 31,612 companies, and 266,547 firm-year observations (Table 1). The average CCC (*IndAdjCCC*) is 82.14 (12.20) days, and the average ROA, *IndAdjROA*, Tobin's *Q*, and *IndAdjTobin's Q* are 0.60%, −1.96%, 1.50, and 0.21, respectively. Japan and the United States exhibit the highest and second highest firm-year observations, accounting for 19.19% and 17.45% of the total sample size, respectively. Different countries exhibit various CCC levels. Greece exhibits the highest CCC, with a mean of 161.75 days, whereas Jordan exhibits the lowest CCC, with a mean of 21.07 days.

In accordance with Fama and French (1997), we also classified firms into 43 industries. Business services, electronic equipment, and retail are the industries with the three highest firm-year observations, accounting for 10.07%, 5.81%, and 5.60% of the total sample size, respectively (Table 2). Shipbuilding, tobacco products, and defense exhibit the lowest firm-year observations, accounting for 0.29%, 0.11%, and 0.08% of the total sample size, respectively. Different industries exhibit various CCC levels. Medical equipment exhibits the highest CCCs, for which ROA, *IndAdjROA*, Tobin's *Q*, and *IndAdjTobin's Q* are −4.16%, −6.07%, 2.31, and 0.43, respectively. Restaurants, hotels, and motels exhibit the lowest CCCs, for which ROA, *IndAdjROA*, Tobin's *Q*, and *IndAdjTobin's Q* are 2.51%, −0.67%, 1.42, and 0.16, respectively.

106.3.1.2 Regression Results for Each Country

Table 3 illustrates the relationship between the CCC and corporate performance for each country. In most countries, the CCC exhibits a negative relationship with firm performance. The CCCs of 40 countries, accounting for 86.96% of the total number of countries, exhibit negative relationships with industry-adjusted ROAs. Among these countries, the *IndAdjCCC* coefficients of 33 countries, accounting for 71.73% of all countries, attain significant levels. Moreover, the findings indicate that the

Table 1 Sample distribution, corporate performance, and CCC by country. This table presents the sample distribution and the mean values of ROA, *IndAdjROA*, Tobin's Q, *IndAdjTobin's Q*, CCC, and *IndAdjCCC* classified by country. ROA is a variable calculated by dividing the net income by the total assets. Tobin's Q is the ratio of the market value of equity added to the book value of debt, divided by the book value of the total assets. The CCC is calculated by adding the number of days of accounts receivable to the number of days of inventory and subtracting the number of days of accounts payable. *IndAdjROA*, *IndAdjTobin's Q*, and *IndAdjCCC* are industry-adjusted ROA, Tobin's Q, and CCC, respectively

Country	Number of firm-years	Percentage (%)	Number of firms	Percentage (%)	ROA	<i>IndAdjROA</i>	Tobin's Q [*]	<i>IndAdj Tobin's Q</i>	CCC	<i>IndAdjCCC</i>
Argentina	492	0.18	57	0.18	0.0418	0.0012	4.5318	1.6428	107.23	2.25
Australia	10,563	3.96	1726	5.46	-0.1353	-0.0958	1.9350	0.4340	91.33	15.42
Austria	904	0.34	114	0.36	0.0207	-0.0025	1.1700	0.0155	49.07	6.13
Belgium	1119	0.42	141	0.45	0.0316	-0.0004	1.4033	0.0922	42.16	5.57
Bermuda	4444	1.67	491	1.55	-0.0077	-0.0332	1.2387	0.2293	110.91	12.95
Brazil	2529	0.95	296	0.94	0.0415	-0.0011	1.1131	0.0612	97.76	3.60
Canada	6099	2.29	801	2.53	-0.0013	-0.0211	1.6601	0.2093	27.94	7.98
Switzerland	2326	0.87	244	0.77	0.0335	-0.0079	1.5163	0.1425	78.10	15.75
Chile	1153	0.43	122	0.39	0.0436	0.0000	1.2004	0.0484	132.57	6.05
China	16,669	6.25	2092	6.62	0.0311	-0.0010	2.1883	0.2551	152.15	23.70
Cayman Islands	2361	0.89	396	1.25	-0.0034	-0.0379	1.7432	0.3739	143.75	12.10
Germany	6830	2.56	875	2.77	-0.0019	-0.0191	1.4403	0.2015	57.46	15.47
Denmark	1445	0.54	180	0.57	0.0232	-0.0088	1.4877	0.1525	78.68	5.76
Spain	1405	0.53	157	0.50	0.0386	-0.0016	1.4211	0.0494	37.58	4.94
Finland	1344	0.50	146	0.46	0.0334	-0.0085	1.4572	0.1047	52.52	6.84
France	7049	2.64	899	2.84	0.0199	-0.0094	1.3923	0.1419	52.03	11.69
	15,994	6.00	2338	7.40	-0.0120	-0.0430	1.6732	0.3010	82.71	10.17

(continued)

Table 1 (continued)

Country	Number of firm-years	Percentage (%)	Number of firms	Percentage (%)	ROA	IndAdjROA	Tobin's Q'	IndAdj Tobin's Q	CCC	IndAdjCCC
United Kingdom										
Greece	1343	0.50	208	0.66	0.0220	-0.0009	1.3144	0.0868	161.75	17.39
Hong Kong	2035	0.76	302	0.96	0.0248	-0.0134	1.1890	0.1162	101.98	14.57
Indonesia	2408	0.90	296	0.94	0.0318	-0.0011	1.2346	0.1171	110.04	12.25
India	13,623	5.11	1790	5.66	0.0524	0.0028	1.3747	0.1966	90.08	16.52
Ireland	741	0.28	95	0.30	0.0103	-0.0112	1.6505	0.0815	61.35	10.05
Israel	1318	0.49	214	0.68	0.0032	-0.0117	1.5216	0.1452	100.87	11.91
Italy	2160	0.81	286	0.90	0.0096	-0.0070	1.2547	0.0761	56.17	5.96
Jordan	459	0.17	92	0.29	0.0424	-0.0001	1.4617	0.0348	21.07	8.66
Japan	51,143	19.19	4233	13.39	0.0135	-0.0035	1.1076	0.0973	62.05	12.83
South Korea	17,473	6.56	1631	5.16	0.0019	-0.0190	1.0276	0.0985	100.92	12.39
Sri Lanka	535	0.20	147	0.47	0.0447	0.0007	1.2801	0.0524	96.10	8.83
Mexico	1112	0.42	123	0.39	0.0414	-0.0014	1.1702	0.0397	74.43	10.85
Malaysia	8106	3.04	960	3.04	0.0196	-0.0104	1.0814	0.1057	123.34	18.43
Netherlands	1906	0.72	226	0.71	0.0344	-0.0082	1.6449	0.1508	50.20	6.95
Norway	1766	0.66	284	0.90	-0.0106	-0.0191	1.5118	0.1265	55.85	6.58

New Zealand	901		0.34	131		0.41	0.0162	-0.0087	1.6319	0.0883	89.42	10.48
Pakistan	1302		0.49	171		0.54	0.0671	0.0041	1.2742	0.0899	105.17	4.33
Peru	517		0.19	60		0.19	0.0757	0.0017	1.2372	0.0348	119.18	7.15
Philippines	1149		0.43	138		0.44	0.0160	-0.0036	1.1992	0.0919	102.31	12.82
Poland	1829		0.69	322		1.02	0.0297	-0.0064	1.4537	0.1350	88.83	9.44
Portugal	525		0.20	72		0.23	0.0154	-0.0008	1.1753	0.0291	56.89	5.12
Russia	718		0.27	120		0.38	0.0744	0.0028	1.5043	0.1364	78.13	6.99
Singapore	5051		1.89	679		2.15	0.0252	-0.0132	1.1888	0.1257	80.27	13.96
Sweden	3365		1.26	488		1.54	-0.0206	-0.0374	1.7608	0.2448	81.96	8.04
Thailand	3990		1.50	455		1.44	0.0412	-0.0050	1.1111	0.0830	104.03	11.31
Turkey	870		0.33	137		0.43	0.0568	0.0004	3.1894	1.0172	88.89	9.42
Taiwan	8679		3.26	1457		4.61	0.0325	-0.0065	1.2903	0.1312	115.26	12.91
United States	46,510		17.45	5088		16.10	-0.0081	-0.0368	1.9996	0.3972	75.42	9.75
South Africa	2287		0.86	332		1.05	0.0665	-0.0068	1.4016	0.0945	44.20	7.54
Sum	266,547		100.00	31,612		100.00						
Mean							0.0060	-0.0196	1.4980	0.2120	82.14	12.20

Table 2 Sample distribution, corporate performance, and CCC by industry. This table presents the sample distribution and the mean values of ROA, *IndAdjROA*, Tobin's *Q*, *IndAdjTobin's Q*, CCC, and *IndAdjCCC* classified by industry. In accordance with Fama and French (1997), we classified firms into 43 industries. ROA is a variable calculated by dividing the net income by the total assets. Tobin's *Q* is the ratio of the market value of equity added to the book value of debt, divided by the book value of the total assets. The CCC is calculated by adding the number of days of accounts receivable to the number of days of inventory and subtracting the number of days of accounts payable. *IndAdjROA*, *IndAdjTobin's Q*, and *IndAdjCCC* are industry-adjusted ROA, Tobin's *Q*, and CCC, respectively

Industry	Number of firm-years	Percentage (%)	Number of firms	Percentage (%)	ROA	<i>IndAdjROA</i>	Tobin's <i>Q</i>	<i>IndAdj Tobin's Q</i>	CCC	<i>IndAdjCCC</i>
Agriculture	1601	0.60	196	0.62	0.0312	-0.0058	1.3550	0.1173	122.38	16.37
Food Products	10,593	3.97	1129	3.57	0.0316	-0.0053	1.2380	0.1120	68.26	10.52
Candy & Soda	797	0.30	82	0.26	0.0313	-0.0039	1.5602	0.1004	48.71	1.06
Beer & Liquor	2315	0.87	256	0.81	0.0360	-0.0062	1.6332	0.0976	139.04	15.91
Tobacco Products	283	0.11	40	0.13	0.0971	0.0010	2.1299	0.0396	100.49	6.15
Recreation	2549	0.96	297	0.94	0.0038	-0.0209	1.3019	0.1472	95.74	11.49
Entertainment	4232	1.59	584	1.85	-0.0181	-0.0259	1.4193	0.1864	55.49	13.81
Printing and Publishing	2842	1.07	322	1.02	0.0264	-0.0131	1.5844	0.1546	71.29	16.57
Consumer Goods	6565	2.46	741	2.34	0.0249	-0.0114	1.6125	0.3022	108.33	14.92
Apparel	5283	1.98	583	1.84	0.0242	-0.0148	1.2613	0.1634	115.50	13.88
Healthcare	2226	0.84	271	0.86	0.0190	-0.0154	1.6707	0.2177	39.02	4.00
Medical Equipment	3933	1.48	488	1.54	-0.0416	-0.0607	2.3132	0.4256	151.25	16.88

Pharmaceutical Products	9558	3.59	1279	4.05	-0.0687	-0.0377	2.4814	0.4205	133.10	21.53
Chemicals	11,619	4.36	1202	3.80	0.0271	-0.0069	1.2692	0.1208	91.16	10.27
Rubber and Plastic Products	4171	1.56	472	1.49	0.0211	-0.0095	1.1929	0.0887	83.02	8.65
Textiles	5177	1.94	581	1.84	0.0076	-0.0083	1.0533	0.0788	115.18	13.10
Construction Materials	11,430	4.29	1210	3.83	0.0240	-0.0076	1.1927	0.1159	101.99	13.65
Construction	10,125	3.80	1077	3.41	0.0157	-0.0072	1.0741	0.0574	109.11	22.25
Steel Works Etc	10,117	3.80	1076	3.40	0.0238	-0.0055	1.1739	0.1116	91.23	9.70
Fabricated Products	1373	0.52	150	0.47	0.0240	-0.0071	1.1913	0.0834	91.30	7.76
Machinery	13,068	4.90	1392	4.40	0.0205	-0.0132	1.4350	0.1965	115.21	17.83
Electrical Equipment	5795	2.17	653	2.07	0.0092	-0.0214	1.4820	0.1991	109.55	10.57
Automobiles and Trucks	7762	2.91	757	2.39	0.0277	-0.0048	1.2453	0.1282	63.87	10.81
Aircraft	821	0.31	71	0.22	0.0334	-0.0016	1.4509	0.0859	123.00	2.66
Shipbuilding	776	0.29	92	0.29	0.0219	-0.0010	1.3629	0.0663	91.46	2.81
Defense	204	0.08	19	0.06	0.0268	-0.0028	1.9473	0.3266	118.86	8.93
Precious Metals	1737	0.65	311	0.98	-0.1313	-0.0832	2.0490	0.3723	89.33	15.05
Mining	3609	1.35	697	2.20	-0.1156	-0.0733	1.9072	0.3636	124.97	20.95
Coal	841	0.32	128	0.40	0.0118	-0.0163	1.9284	0.1568	52.51	5.52
Petroleum and Natural Gas	6889	2.58	1036	3.28	-0.0061	-0.0286	1.5536	0.1835	39.97	6.86

(continued)

Table 2 (continued)

Industry	Number of firm-years	Percentage (%)	Number of firms	Percentage (%)	ROA	IndAdjROA	Tobin's Q*	IndAdj Tobin's Q	CCC	IndAdjCCC
Communication	6027	2.26	880	2.78	-0.0213	-0.0381	1.6299	0.2173	30.97	8.71
Personal Services	2246	0.84	284	0.90	0.0276	-0.0063	1.5662	0.2839	42.92	9.64
Business Services	26,846	10.07	3736	11.82	-0.0172	-0.0431	1.9023	0.4186	54.34	8.71
Computers	8411	3.16	1114	3.52	-0.0201	-0.0386	1.8154	0.3548	81.72	11.68
Electronic Equipment	15,477	5.81	1971	6.23	-0.0076	-0.0327	1.6615	0.2944	102.35	11.68
Measuring Equipment	3154	1.18	348	1.10	0.0045	-0.0264	1.8153	0.2645	143.18	9.52
Business Supplies	5138	1.93	524	1.66	0.0245	-0.0028	1.1792	0.0916	75.68	8.66
Shipping Containers	1816	0.68	184	0.58	0.0302	-0.0018	1.1037	0.0435	76.83	6.80
Transportation	9665	3.63	1081	3.42	0.0270	-0.0060	1.3214	0.1363	28.36	10.90
Wholesale	14,582	5.47	1563	4.94	0.0175	-0.0111	1.1970	0.1338	63.82	12.92
Retail	14,937	5.60	1570	4.97	0.0299	-0.0080	1.4619	0.2235	47.55	14.15
Restaurants, Hotels, Motels	5523	2.07	649	2.05	0.0251	-0.0067	1.4213	0.1608	20.54	5.26
Other	4434	1.66	516	1.63	0.0056	-0.0175	1.3222	0.1907	129.73	12.17
Sum	266,547	100.00	31,612	100.00						
Mean					0.0060	-0.0196	1.4980	0.2120	82.14	12.20

Table 3 The relationship between the CCC and corporate performance for each country.

This table presents the pooled ordinary least squares (OLS) method estimation results for the relationship between the CCC and corporate performance for each country. All regressions include an intercept, country dummies, and year dummies (unreported). The dependent variables are industry-adjusted ROA (*IndAdjROA*) ($\times 10^2$) and industry-adjusted Tobin's Q (*IndAdjTobin's Q*) ($\times 10^2$), respectively, which *IndAdjROA*(*IndAdjTobin's Q*) is the ROA (Tobin's Q) subtracted from the industry median ROA (Tobin's Q) in the corresponding year. ROA is a variable calculated by dividing the net income by the total assets. Tobin's Q is the ratio of the market value of equity added to the book value of debt, divided by the book value of the total assets. The independent variable, the industry-adjusted CCC (*IndAdjCCC*) is calculated by subtracting the CCC from the industry median CCC in the corresponding year. The CCC is calculated by adding the number of days of accounts receivable to the number of days of inventory and subtracting the number of days of accounts payable. Newey–West heteroskedasticity and autocorrelation-robust standard errors are reported in parentheses. ***, **, and * represent 1%, 5%, and 10% significance levels, respectively

Country	<i>IndAdjROA</i>		<i>IndAdjTobin's Q</i>	
	Coefficient	Std. Error	Coefficient	Std. Error
Argentina	-0.0191***	(0.0051)	2.3000**	(0.9790)
Australia	-0.0093***	(0.0013)	0.0041	(0.0038)
Austria	-0.0298**	(0.0141)	0.0182	(0.0221)
Belgium	-0.0031	(0.0061)	0.0028	(0.0489)
Bermuda	-0.0132***	(0.0026)	0.0089	(0.0115)
Brazil	0.0019	(0.0013)	0.0049	(0.0072)
Canada	-0.0079**	(0.0037)	-0.0749***	(0.0262)
Switzerland	-0.0104***	(0.0034)	-0.0154	(0.0357)
Chile	0.0024	(0.0037)	-0.0251**	(0.0107)
China	-0.0036***	(0.0003)	0.0185***	(0.0055)
Cayman Islands	-0.0209***	(0.0037)	-0.0146	(0.0181)
Germany	-0.0038	(0.0024)	-0.0025	(0.0194)
Denmark	-0.0072	(0.0060)	-0.0735**	(0.0311)
Spain	-0.0030*	(0.0017)	-0.0426**	(0.0186)
Finland	-0.0053	(0.0047)	-0.0184	(0.0304)
France	-0.0080***	(0.0017)	-0.0192*	(0.0099)
United Kingdom	-0.0160***	(0.0016)	0.0192*	(0.0115)
Greece	-0.0051***	(0.0011)	-0.0761***	(0.0101)
Hong Kong	-0.0093***	(0.0026)	0.0002	(0.0080)
Indonesia	-0.0105***	(0.0019)	-0.0441**	(0.0185)
India	-0.0073***	(0.0006)	-0.0904***	(0.0064)
Ireland	-0.0091	(0.0063)	-0.0650	(0.0774)
Israel	-0.0240***	(0.0056)	-0.0889***	(0.0331)
Italy	-0.0114***	(0.0029)	-0.0234	(0.0177)
Jordan	0.0214***	(0.0078)	0.1130*	(0.0626)
Japan	-0.0067***	(0.0004)	-0.0425***	(0.0044)
South Korea	-0.0252***	(0.0019)	-0.0175***	(0.0067)
Sri Lanka	-0.0085**	(0.0041)	-0.0159	(0.0352)
Mexico	-0.0014	(0.0026)	-0.0640***	(0.0183)
Malaysia	-0.0095***	(0.0009)	-0.0314***	(0.0035)
Netherlands	-0.0011	(0.0043)	0.0800	(0.0685)

(continued)

Table 3 (continued)

Country	<i>IndAdjROA</i>		<i>IndAdjTobin's Q</i>	
	Coefficient	Std. Error	Coefficient	Std. Error
Norway	−0.0104**	(0.0042)	0.0056	(0.0143)
New Zealand	0.0002	(0.0007)	−0.0044**	(0.0019)
Pakistan	−0.0076*	(0.0044)	−0.0245	(0.0320)
Peru	−0.0040	(0.0066)	−0.0766	(0.0538)
Philippines	−0.0020*	(0.0011)	0.0026	(0.0132)
Poland	−0.0084**	(0.0035)	−0.0215	(0.0355)
Portugal	−0.0059***	(0.0014)	−0.0474***	(0.0082)
Russia	0.0014	(0.0034)	−0.0219	(0.0681)
Singapore	−0.0108***	(0.0024)	−0.0288**	(0.0113)
Sweden	−0.0361***	(0.0059)	0.0373	(0.0333)
Thailand	−0.0187***	(0.0019)	−0.0767***	(0.0095)
Turkey	−0.0016	(0.0027)	0.1930	(0.4390)
Taiwan	−0.0088***	(0.0010)	−0.0261***	(0.0037)
United States	0.0052***	(0.0019)	−0.2370***	(0.0147)
South Africa	−0.0087**	(0.0038)	−0.0858***	(0.0184)
Aggregate coefficient estimates	−0.0105***	(0.0004)	−0.0347***	(0.0023)

CCCs of 31 countries, accounting for 67.39% of all countries, exhibit negative relationships with industry-adjusted Tobin's Q . Among these countries, the *IndAdjCCC* coefficients of 20 countries, accounting for 43.48% of all countries, attain significant levels. Among all countries, the CCC of Sweden exhibits the most significant effect on the *IndAdjROA*, for which the coefficient of *IndAdjCCC* is -0.0361 at a 1% significance level. The CCC of the United States exhibits the most significant negative effect on *IndAdjTobin's Q*, for which the coefficient of *IndAdjCCC* is -0.2370 at a 1% significance level.

106.3.1.3 Regression Results for Each Industry

Table 4 illustrates the relationship between the CCC and corporate performance for each industry. In most industries, the CCC exhibits a negative relationship with firm performance. The CCCs of 40 industries, accounting for 93.02% of the total number of industries, exhibit negative relationships with *IndAdjROA*. Among these industries, the *IndAdjCCC* coefficients of 36 industries, accounting for 83.72% of all industries, attain significant levels. Moreover, the results indicate that the CCCs of 31 industries, accounting for 72.09% of all industries, exhibit negative relationships with industry-adjusted Tobin's Q . Among these industries, the *IndAdjCCC* coefficients of 19 industries, accounting for 44.19% of all industries, attain significant levels. Among the various industries, the CCC of the candy and soda industry exhibits the most significant effect on the *IndAdjROA*, for which the coefficient of *IndAdjCCC* is -0.0363 at a 1% significance level. The CCC of personal services exhibits the most significant effect on *IndAdjTobin's Q*, for which the coefficient of *IndAdjCCC* is -0.1630 at a 1% significance level.

Table 4 The relationship between the CCC and corporate performance for each industry.

This table presents the pooled ordinary least squares (OLS) method estimation results for the relationship between the CCC and corporate performance for each industry. All regressions include an intercept, industry dummies, and year dummies (unreported). The dependent variables are industry-adjusted ROA (*IndAdjROA*) ($\times 10^2$) and industry-adjusted Tobin's Q (*IndAdjTobin's Q*) ($\times 10^2$), respectively, which *IndAdjROA*(*IndAdjTobin's Q*) is the ROA (Tobin's Q) subtracted from the industry median ROA (Tobin's Q) in the corresponding year. ROA is a variable calculated by dividing the net income by the total assets. Tobin's Q is the ratio of the market value of equity added to the book value of debt, divided by the book value of the total assets. The independent variable, the industry-adjusted CCC (*IndAdjCCC*) is calculated by subtracting the CCC from the industry median CCC in the corresponding year. The CCC is calculated by adding the number of days of accounts receivable to the number of days of inventory and subtracting the number of days of accounts payable. Newey–West heteroskedasticity and autocorrelation-robust standard errors are reported in parentheses. ***, **, and * represent 1%, 5%, and 10% significance levels, respectively

Industry	<i>IndAdjROA</i>		<i>IndAdjTobin's Q</i>	
	Coefficient	Std. Error	Coefficient	Std. Error
Agriculture	-0.0044*	(0.0024)	0.0392***	(0.0151)
Food Products	-0.0086***	(0.0012)	-0.0722***	(0.0094)
Candy & Soda	-0.0363***	(0.0138)	0.0579	(0.0512)
Beer & Liquor	-0.0045**	(0.0018)	-0.0244	(0.0151)
Tobacco Products	0.0019	(0.0061)	-0.0339	(0.0524)
Recreation	-0.0168***	(0.0038)	-0.0556***	(0.0207)
Entertainment	-0.0120***	(0.0029)	-0.0255*	(0.0140)
Printing and Publishing	-0.0136***	(0.0032)	-0.0910***	(0.0231)
Consumer Goods	-0.0119***	(0.0017)	-0.0891***	(0.0119)
Apparel	-0.0046***	(0.0018)	-0.0128	(0.0125)
Healthcare	-0.0291***	(0.0066)	-0.0886*	(0.0456)
Medical Equipment	0.0003	(0.0034)	0.0035	(0.0214)
Pharmaceutical Products	0.0109***	(0.0019)	-0.0057	(0.0125)
Chemicals	-0.0144***	(0.0019)	0.0316***	(0.0111)
Rubber and Plastic Products	-0.0117***	(0.0028)	-0.0126	(0.0117)
Textiles	-0.0089***	(0.0018)	-0.0009	(0.0173)
Construction Materials	-0.0103***	(0.0011)	-0.0407***	(0.0072)
Construction	-0.0055***	(0.0007)	-0.0337***	(0.0038)
Steel Works Etc	-0.0051***	(0.0017)	0.0406***	(0.0135)
Fabricated Products	-0.0109***	(0.0038)	-0.0317	(0.0291)
Machinery	-0.0165***	(0.0013)	-0.0133	(0.0084)
Electrical Equipment	-0.0166***	(0.0022)	-0.0751***	(0.0165)
Automobiles and Trucks	-0.0085***	(0.0018)	-0.0070	(0.0116)
Aircraft	-0.0046	(0.0037)	-0.0643**	(0.0311)
Shipbuilding	-0.0065	(0.0056)	-0.0152	(0.0270)
Defense	-0.0075	(0.0071)	0.4010***	(0.1250)
Precious Metals	-0.0206***	(0.0051)	0.0407	(0.0289)
Mining	-0.0114***	(0.0026)	-0.0229	(0.0170)
Coal	-0.0106*	(0.0059)	0.1210	(0.0769)
Petroleum and Natural Gas	-0.0152***	(0.0026)	0.0190	(0.0145)
Communication	-0.0168***	(0.0031)	-0.0901***	(0.0156)

(continued)

Table 4 (continued)

Industry	<i>IndAdjROA</i>		<i>IndAdjTobin's Q</i>	
	Coefficient	Std. Error	Coefficient	Std. Error
Personal Services	−0.0091**	(0.0040)	−0.1630***	(0.0367)
Business Services	−0.0164***	(0.0015)	−0.1230***	(0.0096)
Computers	−0.0311***	(0.0027)	−0.0641***	(0.0150)
Electronic Equipment	−0.0261***	(0.0019)	−0.0506***	(0.0115)
Measuring Equipment	−0.0203***	(0.0038)	−0.0484**	(0.0225)
Business Supplies	−0.0094***	(0.0028)	−0.0114	(0.0163)
Shipping Containers	−0.0018	(0.0024)	0.0044	(0.0253)
Transportation	−0.0077***	(0.0016)	0.0028	(0.0124)
Wholesale	−0.0085***	(0.0014)	−0.0222**	(0.0102)
Retail	−0.0056***	(0.0012)	−0.0850***	(0.0110)
Restaurants, Hotels, Motels	−0.0030***	(0.0011)	−0.0883***	(0.0074)
Other	−0.0056***	(0.0017)	0.0012	(0.0090)
Aggregate coefficient estimates	−0.0105***	(0.0004)	−0.0347***	(0.0023)

106.3.1.4 Differences in Corporate Performance Between High- and Low-CCC Firms

Table 5 illustrates the differences in corporate performance between firms that implement aggressive policies and those that implement conservative policies for WCM. We classify firms into two groups based on the median CCC (*IndAdjCCC*): low-CCC (*low-IndAdjCCC*) firms (below the median CCC [*IndAdjCCC*]; the aggressive liquidity policy group), and high-CCC (*high-IndAdjCCC*) firms (above the median of CCC [*IndAdjCCC*]; the conservative liquidity policy group). The mean and median variations are assessed using the t-test and the Wilcoxon rank-sum test. Based on various performance indicators, low-CCC (*low-IndAdjCCC*) firms exhibit higher mean and median values, both of which attain significant levels. The results indicate that firms with lower CCCs exhibit higher corporate performance; for example, the difference in the *IndAdjROA* mean (median) between *low-IndAdjCCC* firms and *high-IndAdjCCC* firms is 0.0096 (0.0017), both of which are statistically significant at a 1% significance level. The difference in *IndAdjTobin's Q* mean (median) between *low-IndAdjCCC* firms and *high-IndAdjCCC* firms is 0.0675 (0.0272), both of which are statistically significant at a 1% significance level.

106.3.2 Regression Results

After controlling for industry-fixed effects, country-fixed effects, year dummies, and related control variables, the results indicate that the *IndAdjCCC* for all regression models exhibits significantly negative relationships with *IndAdjROA* and *IndAdjTobin's Q* at a 5% significance level or better (Table 6). Therefore, firms can shorten their CCCs to increase profitability and value. These findings support that an aggressive liquidity policy can enhance a firm's operating performance and value. By contrast, a

Table 5 Difference in corporate performance between high- and low-CCC firms. This table presents the differences in corporate performance between high- and low-CCC firms. We compare the difference in the mean and median between the two samples based on the median CCC (*IndAdjCCC*): low-CCC/low-*IndAdjCCC* firms (below the median of *CCC/IndAdjCCC*), and high-CCC/high-*IndAdjCCC* firms (above the median of *CCC/IndAdjCCC*). ROA is a variable calculated by dividing the net income by the total assets. Tobin's Q is the ratio of the market value of equity added to the book value of debt, divided by the book value of the total assets. The CCC is calculated by adding the number of days of accounts receivable to the number of days of inventory and subtracting the number of days of accounts payable. *IndAdjROA*, *IndAdjTobin's Q*, and *IndAdjCCC* are industry-adjusted ROA, Tobin's Q, and CCC, respectively. Differences in the mean and median are assessed using the t-test and the Wilcoxon rank-sum test. ***, **, and * represent 1%, 5%, and 10% significance levels, respectively

Panel A. The sample firms are classified into two groups based on CCC				
		Low CCC firms	High CCC firms	The difference of mean and median
<i>IndAdjROA</i>	Mean	0.0023	-0.0022	0.0045***
	Median	0.0076	0.0067	0.0010***
<i>IndAdjTobin's Q</i>	Mean	0.0357	-0.0342	0.0699***
	Median	-0.0849	-0.1189	0.0340***
Panel B. The sample firms are classified into two groups based on <i>IndAdjCCC</i>				
		Low <i>IndAdjCCC</i> firms	High <i>IndAdjCCC</i> firms	The difference of mean and median
<i>IndAdjROA</i>	Mean	0.0042	-0.0055	0.0096***
	Median	0.0079	0.0062	0.0017***
<i>IndAdjTobin's Q</i>	Mean	0.0292	-0.0383	0.0675***
	Median	-0.0943	-0.1215	0.0272***

conservative WCM policy can harm a firm's performance. Moreover, the results indicate that the interaction term (*IndAdjCCC* $\times$ *LowCCC*) is significantly positive at a 1% significance level, suggesting that the negative relationships between CCC and the firm's performance diminish or reverse when the industry-adjusted CCC is below 0. This finding indicates that firms can shorten their CCC to increase profitability and value; however, this effect reduces or reverses when firms exist at the lower CCC level. The sum of the coefficients of the interaction term (*IndAdjCCC* $\times$ *LowCCC*) and *IndAdjCCC* for *IndAdjTobin's Q* is positive. From this result, we can infer that CCC has a significantly positive relationship with firm value for extremely low CCC firms.³

Table 6 also indicates that *Size* exhibits significantly positive relationships with *IndAdjROA* and *IndAdjTobin's Q*, implying that larger firms exhibit higher

³ We also test whether the relation between ROA (Tobin's Q) and CCC for low CCC and high CCC firms is different. We divide the sample countries into two groups based on the CCC industry median. The results indicate that the *IndAdjCCC* for high CCC firms (i.e., value is above industry median) exhibit significantly negative relationships with *IndAdjROA* and *IndAdjTobin's Q* at a 5% significance level or better. However, the results show that the *IndAdjCCC* for low CCC firms (i.e., value is below industry median) exhibit significantly positive (negative) relationships with *IndAdjTobin's Q* (*IndAdjROA*). These findings support the results of the interaction term.

Table 6 The relationship between the CCC and corporate performance for all countries. This table presents the pooled ordinary least squares (OLS) method estimation results for the relationship between the CCC and corporate performance for all countries. All regressions include an intercept, industry dummies, country dummies, and year dummies (unreported). The dependent variables are industry-adjusted ROA (*IndAdjROA*) ($\times 10^2$) and industry-adjusted Tobin's Q (*IndAdjTobin's Q*) ($\times 10^2$), respectively, which *IndAdjROA*(*IndAdjTobin's Q*) is the ROA (Tobin's Q) subtracted from the industry median ROA (Tobin's Q) in the corresponding year. ROA is a variable calculated by dividing the net income by the total assets. Tobin's Q is the ratio of the market value of equity added to the book value of debt, divided by the book value of the total assets. The independent variable, the industry-adjusted CCC (*IndAdjCCC*) is calculated by subtracting the CCC from the industry median CCC in the corresponding year. The CCC is calculated by adding the number of days of accounts receivable to the number of days of inventory and subtracting the number of days of accounts payable. The interaction term (*IndAdjCCC* $\times$ *LowCCC*) is calculated by multiplying the industry-adjusted CCC dummy (*LowCCC*) and industry-adjusted CCC (*IndAdjCCC*). The industry-adjusted CCC dummy variable (*LowCCC*) equals 1, if the industry-adjusted CCC is negative and 0 otherwise. Firm size (*Size*) is defined as the natural logarithm of the market value of equity. Payout ratio (*DIV*) is defined as the ratio of dividends divided by the operating revenues. *CAPEXP* denotes the ratio of capital expenditure and other investments divided by the total assets. Leverage (*LEV*) is defined as the ratio of the total debt divided by the total assets. *LagROA* denotes the ROA of the previous year. *RDR* is the ratio of research and development expenditure divided by the total assets. *STDROA* is the standard deviation of the ROA over the preceding 5-year period. *MB* denotes the ratio of the market value of equity divided by the book value of equity. Newey–West heteroskedasticity and autocorrelation-robust standard errors are reported in parentheses. ***, **, and * represent 1%, 5%, and 10% significance levels, respectively

	<i>IndAdjROA</i>	<i>IndAdjROA</i>	<i>IndAdjTobin's Q</i>	<i>IndAdjTobin's Q</i>
<i>IndAdjCCC</i>	−0.0085*** (0.0003)	−0.0098*** (0.0004)	−0.0341*** (0.0022)	−0.0058** (0.0028)
<i>IndAdjCCC</i> × <i>LowCCC</i>		0.0049*** (0.0010)		0.0716*** (0.0061)
<i>Size</i>	0.9750*** (0.0266)	0.9730*** (0.0264)	10.8980*** (0.0918)	10.0570*** (0.0853)
<i>DIV</i>	16.0170*** (0.7460)	16.0180*** (0.7450)	59.0740*** (3.8590)	60.8150*** (3.6610)
<i>CAPEXP</i>	1.7490*** (0.3590)	1.7740*** (0.3600)	57.5970*** (2.2680)	54.9550*** (2.1060)
<i>LEV</i>	−7.4180*** (0.1610)	−7.4070*** (0.1610)	−26.6160*** (0.8370)	−19.1710*** (0.7770)
<i>LagROA</i>	14.0970*** (1.4770)	14.0820*** (1.4760)	9.2540*** (2.4880)	11.2630*** (2.3080)
<i>RDR</i>	−41.7310*** (1.4550)	−41.6630*** (1.4490)	209.0330*** (6.5160)	170.0100*** (5.9810)
<i>STDROA</i>	−1.8880*** (0.6380)	−1.8690*** (0.6380)	47.1500*** (3.2690)	40.4110*** (3.1190)
<i>MB</i>	0.2200*** (0.0231)	0.2210*** (0.0231)		

(continued)

Table 6 (continued)

	<i>IndAdjROA</i>	<i>IndAdjROA</i>	<i>IndAdjTobin's Q</i>	<i>IndAdjTobin's Q</i>
<i>Industry dummies</i>	Yes	Yes	Yes	Yes
<i>Country dummies</i>	Yes	Yes	Yes	Yes
<i>Year dummies</i>	Yes	Yes	Yes	Yes
<i>Adjusted R²</i>	0.2639	0.2640	0.1309	0.1170
<i>F-value</i>	689.27***	683.65***	292.70***	252.16***

performance. *DIV* exhibits significantly positive relationships with the two types of firm performance variables. The coefficients of *CAPEXP* (*LagROA*) for *IndAdjROA* and *IndAdjTobin's Q* are significantly positive at a 1% significance level, suggesting that higher capital expenditure (previous ROA) can increase a firm's performance. *LEV* exhibits a significantly negative association with performance measures, denoting that an increase in financial leverage may reduce a firm's performance. *STDROA* exhibits a significantly negative relationship with *IndAdjROA*, but a positive relationship with *IndAdjTobin's Q*. *RDR* exhibits a negative relationship with *IndAdjROA* but is expected to increase a firm's value. Thus, *RDR* exhibits a positive relationship with *IndAdjTobin's Q*. The coefficients of *MB* for *IndAdjROA* are positive and statistically significant at a 1% significance level, implying that growth firms (i.e., high *MB* firms) achieve high profitability.

106.4 Endogeneity

According to Petersen and Rajan (1997), Shin and Soenen (1998), Opler et al. (1999), Wang (2002), Chiou et al. (2006), Bates et al. (2009), and Baños-Caballero et al. (2010), a firm's profitability and value can influence the WCM. We adopt the following two approaches to address endogeneity problems: a three-stage least squares (3SLS) method and the generalized method of moments (GMM).

106.4.1 Three-Stage Least Squares

To control for the potential effects of profitability and value on the CCC, we estimate two pairs of equations simultaneously using a 3SLS procedure: Eqs. (3) and (4) and Eqs. (5) and (6):

$$\begin{aligned}
IndAdjROA_{i,t} = & \beta_0 + \beta_1 IndAdjCCC_{i,t} + \beta_2 IndAdjCCC_{i,t} \times LowCCC_{i,t} \\
& + \beta_3 SIZE_{i,t} + \beta_4 DIV_{i,t} + \beta_5 CAPEXP_{i,t} + \beta_6 LEV_{i,t} + \beta_7 LagROA_{i,t} \\
& + \beta_8 RDR_{i,t} + \beta_9 STDROA_{i,t} + \beta_{10} MB_{i,t} + Industry\ dummies \\
& + Country\ dummies + Year\ dummies + \varepsilon_{i,t}
\end{aligned}
\tag{3}$$

$$\begin{aligned}
IndAdjCCC_{i,t} = & \beta_0 + \beta_1 IndAdjROA_{i,t} + \beta_2 SIZE_{i,t} + \beta_3 LEV_{i,t} + \beta_4 SG_{i,t} \\
& + \beta_5 STDSALES_{i,t} + \beta_6 CF_{i,t} + \beta_7 FA_{i,t} + \beta_8 DISTRESS_{i,t} \\
& + Industry\ dummies + Country\ dummies + Year\ dummies + \varepsilon_{i,t}
\end{aligned}
\tag{4}$$

$$\begin{aligned}
IndAdjTobin's\ Q_{i,t} = & \beta_0 + \beta_1 IndAdjCCC_{i,t} + \beta_2 IndAdjCCC_{i,t} \times LowCCC_{i,t} \\
& + \beta_3 SIZE_{i,t} + \beta_4 DIV_{i,t} + \beta_5 CAPEXP_{i,t} + \beta_6 LEV_{i,t} + \beta_7 LagROA_{i,t} \\
& + \beta_8 RDR_{i,t} + \beta_9 STDROA_{i,t} + Industry\ dummies \\
& + Country\ dummies + Year\ dummies + \varepsilon_{i,t}
\end{aligned}
\tag{5}$$

$$\begin{aligned}
IndAdjCCC_{i,t} = & \beta_0 + \beta_1 IndAdjTobin's\ Q_{i,t} + \beta_2 SIZE_{i,t} + \beta_3 LEV_{i,t} \\
& + \beta_4 SG_{i,t} + \beta_5 STDSALES_{i,t} + \beta_6 CF_{i,t} + \beta_7 FA_{i,t} \\
& + \beta_8 DISTRESS_{i,t} + Industry\ dummies + Country\ dummies \\
& + Year\ dummies + \varepsilon_{i,t}
\end{aligned}
\tag{6}$$

In accordance with Myers and Majluf (1984), Emery (1987), Whited (1992), Fazzari and Petersen (1993), Petersen and Rajan (1997), Chiou et al. (2006), Kieschnich et al. (2006), Cuñat (2007), Uyar (2009), Molina and Preve (2009), Baños-Caballero et al. (2010), and Hill et al. (2010), we set the control variables in Eqs. (4) and (6) to include *Size*, *SG*, *LEV*, *STDSALES*, *CF*, *FA*, *DISTRESS*, *Industry dummies*, *Country dummies*, and *Year dummies*. *Size* is defined as the natural logarithm of the market value of equity; *SG* represents the percentage changes in operating revenues in the previous year. *STDSALES* represents the standard deviation of operating revenue over the preceding 5-year period. *LEV* is defined as the ratio of the total debt divided by the total assets. *CF* represents the ratio of the net income added to depreciation divided by the total assets. *FA* is calculated as the ratio of tangible fixed assets to total assets. *DISTRESS* is equal to 1 if a firm fulfills the definition of financial distress, proposed by Molina and Preve (2009) and is 0 otherwise.⁴

⁴In accordance with Molina and Preve (2009), a firm must satisfy two criteria to be classified as financially distressed. First, a coverage ratio is calculated as the operating income before depreciation divided by an interest expense of less than one for 2 consecutive years or less than 0.80 in any given year. Second, a firm is considered overleveraged if its leverage ratio is in the top two deciles of the leverage ratio of its industry in a given year.

Table 7 The relationship between corporate performance and the CCC: 3SLS method of estimation. This table presents the 3SLS estimation results for the relationship between the CCC and corporate performance for all countries. All regressions include an intercept, industry dummies, country dummies, and year dummies (unreported). The dependent variables are industry-adjusted ROA (*IndAdjROA*) and industry-adjusted Tobin's Q (*IndAdjTobin's Q*), respectively, which *IndAdjROA(IndAdjTobin's Q)* is the ROA (Tobin's Q) subtracted from the industry median ROA (Tobin's Q) in the corresponding year. ROA is a variable calculated by dividing the net income by the total assets. Tobin's Q is the ratio of the market value of equity added to the book value of debt, divided by the book value of the total assets. The independent variable, the industry-adjusted CCC (*IndAdjCCC*) is calculated by subtracting the CCC from the industry median CCC in the corresponding year. The CCC is calculated by adding the number of days of accounts receivable to the number of days of inventory and subtracting the number of days of accounts payable. The interaction term (*IndAdjCCC* $\times$ *LowCCC*) is calculated by multiplying the industry-adjusted CCC dummy (*LowCCC*) and industry-adjusted CCC (*IndAdjCCC*). The industry-adjusted CCC dummy variable (*LowCCC*) equals 1 if the industry-adjusted CCC is negative and 0 otherwise. *Size*, *DIV*, *CAPEXP*, *LEV*, *LagROA*, *RDR*, *STDROA*, and *MB* are defined in Table 6. *SG* represents the percentage changes in operating revenues in the previous year. *STDSALES* represents the standard deviation of operating revenues over the preceding 5-year period. *CF* represents the ratio of the net income added to depreciation divided by the total assets. *FA* is calculated as the ratio of tangible fixed assets to total assets. *DISTRESS* is equal to 1 if a firm fulfills the definition of financial distress proposed by Molina and Preve (2009) and is 0 otherwise. Standard errors are reported in parentheses. ***, **, and * represent 1%, 5%, and 10% significance levels, respectively

	<i>IndAdjROA</i>	<i>IndAdjCCC</i>	<i>IndAdjTobin's Q</i>	<i>IndAdjCCC</i>
<i>IndAdjCCC</i>	−0.0145*** (0.0006)		−0.0352*** (0.0020)	
<i>IndAdjCCC</i> $\times$ <i>LowCCC</i>	0.0046*** (0.0023)		0.1539*** (0.0079)	
<i>IndAdjROA</i>		496.8657*** (35.8972)		
<i>IndAdjTobin's Q</i>				0.5867 (0.4297)
<i>Size</i>	−0.0149*** (0.0016)	−1.1944*** (0.1150)	0.0322*** (0.0056)	−0.1778* (0.1072)
<i>DIV</i>	0.1351*** (0.0426)		0.7365*** (0.1481)	
<i>CAPEXP</i>	0.0816** (0.0333)		1.3734*** (0.1134)	
<i>LEV</i>	−0.0545*** (0.0134)	−23.5188*** (0.9498)	0.4357*** (0.0522)	−31.7160*** (0.8296)
<i>LagROA</i>	0.0007 (0.0146)		−0.3203*** (0.0508)	
<i>RDR</i>	0.2899*** (0.0636)		3.2506*** (0.2092)	
<i>STDROA</i>	0.1593*** (0.0166)		0.9085*** (0.0564)	
<i>MB</i>	0.0128*** (0.0012)			

(continued)

Table 7 (continued)

	<i>IndAdjROA</i>	<i>IndAdjCCC</i>	<i>IndAdjTobin's Q</i>	<i>IndAdjCCC</i>
<i>SG</i>		−20.7433*** (0.4359)		−17.8612*** (0.4103)
<i>STDSALES</i>		−0.0018*** (0.0002)		−0.0020*** (0.0002)
<i>CF</i>		−436.1760*** (27.9255)		−51.2648*** (1.6591)
<i>FA</i>		−7.1388*** (1.7444)		−33.0440*** (0.8456)
<i>DISTRESS</i>		9.3065*** (0.8297)		4.1548*** (0.6998)
<i>Industry dummies</i>	Yes	Yes	Yes	Yes
<i>Country dummies</i>	Yes	Yes	Yes	Yes
<i>Year dummies</i>	Yes	Yes	Yes	Yes
<i>Adjusted R²</i>	0.0051	0.0351	0.0059	0.0401
<i>F-value</i>	10.68***	71.17***	12.15***	80.43***

The results of 3SLS estimation indicate that *IndAdjCCC* continues to exhibit significantly negative relationships with *IndAdjROA* and *IndAdjTobin's Q* at a 1% significance level, and the coefficient of the interaction term *IndAdjCCC* $\times$ *LowCCC* remains significantly positive (Table 7). These results support that aggressive WCM policy can enhance corporate performance.⁵ However, the negative relationships between CCC and the firm's performance diminish or reverse when the industry-adjusted CCC is below 0. Column (2) also indicates that *IndAdjROA* exhibits a significantly positive relationship with *IndAdjCCC*, suggesting that firms with higher returns on assets have higher CCCs.

106.4.2 Generalized Method of Moments

To robustly avoid endogeneity problems, we apply the GMM methodology of Arellano and Bond (1991). They suggest applying the differences as the first step and using suitable lagged levels of dependent variables as instruments to control endogeneity as the second step. The results indicate that the *IndAdjCCC* (*IndAdjCCC* $\times$ *LowCCC*) exhibits significantly negative (positive) relationships with the two types of firm performance variables at a 1% significance level. These

⁵We also perform regressions for dependent variable ROA and Tobin's Q (independent variable CCC), and obtain similar results. We do not tabulate the results of dependent variable ROA and Tobin's Q (independent variable CCC).

findings imply that a shortened working cycle can increase firm performance, and this effect reduces or reverses when firms exist at the lower CCC level.⁶

106.5 Robustness Checks

106.5.1 Macroeconomic Environment

Existing literature indicates that macroeconomic changes influence the WCM policy and liquidity. Michaelas et al. (1999) suggest that small businesses rely more heavily on short-term financing, rendering them more sensitive to macroeconomic changes. Smith (1987) argues that the state of the economy influences the level of accounts receivable. Blinder and Maccini (1991) observe that recessions are related to severe inventory reductions. Hence, the influence of the CCC on a firm's performance may differ under the circumstances of economic boom or economic recession. In accordance with Blinder and Maccini (1991), Michaelas et al. (1999), and Booth et al. (2001), we adopt *GDPG* and *INFLATION* to divide the sample firms into two groups based on the macroeconomic variable median for each year (above/below the median), namely, the high *GDPG* (*INFLATION*) group and the low *GDPG* (*INFLATION*) group. We subsequently rerun Eqs. (1) and (2) to control the influence of macroeconomic changes on CCC. *GDPG* denotes the annual growth rate of real per capita GDP. *INFLATION* is the annual growth rate of the consumer price index.

After controlling changes in macroeconomic factors, our conclusion remains unchanged. *IndAdjCCC* (*IndAdjCCC* $\times$ *LowCCC*) exhibits significantly negative (positive) relationships with *IndAdjROA* and *IndAdjTobin's Q* in both high and low *GDPG* groups. Moreover, *IndAdjCCC* (*IndAdjCCC* $\times$ *LowCCC*) exhibits a significantly negative (positive) relationship with *IndAdjROA* in both high and low *INFLATION* groups. The results support that an aggressive operating WCM policy can increase a firm's performance, and this effect diminishes or reverses when firms exist at the lower CCC level.⁷

106.5.2 Economic Development Status

Klapper (2006) observes that the economic development status influences a business' accounts receivable by changing the credit policy. Céspedes et al. (2010) indicate that debt markets in Latin American are small and inefficient, allowing firms only limited debt options such as long-term bonds. Moreover, debt costs are

⁶We also perform regressions for dependent variable ROA and Tobin's Q (independent variable CCC), and obtain similar results. We do not tabulate the results of dependent variable ROA and Tobin's Q (independent variable CCC).

⁷We also perform regressions for dependent variable ROA and Tobin's Q (independent variable CCC), and obtain similar results. In the described robustness checks, to save space, we do not tabulate the results of dependent variable ROA and Tobin's Q (independent variable CCC).

high for average firms in the region. Consequently, the influence of the CCC on firm performance can vary between developed and developing economies.

To control this scenario, we divide the sample countries into two groups: developed economies and developing economies and rerun Eqs. (1) and (2). The economic development status (developed economies versus developing economies) is classified according to the World Bank.⁸ The results illustrate the relationship between *IndAdjCCC* and a firm's performance for both developed and developing economies. For both, the coefficients of *IndAdjCCC* for *IndAdjROA* and *IndAdjTobin's Q* remain significantly negative, indicating that the CCCs exhibit negative relationships with profitability and value.⁹ Moreover, the results indicate that the interaction term (*IndAdjCCC* $\times$ *LowCCC*) is significantly positive at the 1% level, suggesting that the negative relationships between CCC and the firm's performance diminish or reverse when the industry-adjusted CCC is below 0.

106.5.3 Financial Crises

Campello et al. (2010) surveyed 1050 Chief Financial Officers in 39 countries in North America, Europe, and Asia to directly assess whether the officers' respective firms were credit-constrained during the global financial crisis of 2008. They observed that during the crisis, financially constrained firms planned to cut investment, technology, marketing, and employment at a higher rate than financially unconstrained firms did. They also indicate that constrained firms were forced to employ a sizeable portion of their cash savings during the crisis, and to significantly cut their planned dividend distributions.

The influence of CCCs on firm performance can be different between financial crisis and nonfinancial crisis periods. We divide the sample countries into two groups: financial crisis period and nonfinancial crisis period and rerun Eqs. (1) and (2). The financial crisis period denotes the period during which a country experiences a banking or currency crisis, for which the dates are provided by Reinhart and Rogoff (2011). The results still support that an aggressive WCM policy can increase firm performance; however, this effect reduces or reverses when firms exist at the lower CCC level.¹⁰

⁸We also use alternative classifications from the Human Development Index of the United Nations Development Program and the International Monetary Fund. The results are similar.

⁹We also perform regressions for dependent variable ROA and Tobin's Q (independent variable CCC), and obtain similar results. In the described robustness checks, to save space, we do not tabulate the results of dependent variable ROA and Tobin's Q (independent variable CCC).

¹⁰We also perform regressions for dependent variable ROA and Tobin's Q (independent variable CCC), and obtain similar results. In the described robustness checks, to save space, we do not tabulate the results of dependent variable ROA and Tobin's Q (independent variable CCC).

106.5.4 Corporate Governance

Hail and Leuz (2006) observe that firms in countries with strong legal protection for investors tend to enjoy lower equity costs than firms in countries with weak legal protection for investors do. Chen et al. (2009) documented that firms with strong firm-level corporate governance have lower capital costs, particularly those in countries with weak legal protection. Shleifer and Wolfenzon (2002), Almeida et al. (2011), Kusunadi and Wei (2011), and Kuan et al. (2012) all indicate that corporate governance influences capital costs and the changes in a firm's cash management policy.

To control for the effects of divergence in corporate governance on the relationship between liquidity and firm performance, we divided the sample countries into two groups based on the corporate governance variable median (above and below the median): the high level of anti-self-dealing (anti-director) index group and the low level of anti-self-dealing (anti-director) index group. We rerun Eqs. (1) and (2). The anti-self-dealing and anti-director exhibit high numbers, and both indicate strong investor protection. The anti-self-dealing index and the anti-director index were constructed by Djankov et al. (2008). These indices measure minority shareholder protection against the actions of the controlling shareholder that may hurt the shareholder value at the country level. The results provide evidence of the negative relationship between the CCC and firm performance classified by corporate governance, indicating that firms can shorten their CCC to improve performance.¹¹ However, the negative relationships between CCC and the firm's performance diminish or reverse when the industry-adjusted CCC is below 0.

106.5.5 Financial Constraints

Riddiough and Wu (2009) identify substantial differences between the investment and liquidity management policies of firms and found that more (less) financially constrained firms exhibit high (low) investment and liquidity management sensitivity to variables that are measures of financial market friction. Ang and Smedema (2011) observe that firms do not always prepare for future recession because of financial constraints and low quantities of cash.

To control the financially constrained effect on the relationship between the CCC and firm performance, we divided the sample firms into two groups: the financially constrained group and the financially unconstrained group and rerun Eqs. (1) and (2). Luo (2011) and Lin et al. (2013) suggest that larger firms are generally viewed as less financially constrained than smaller ones are. Fazzari et al. (1988) and Tsai (2014) argue that unconstrained firms are more likely to exhibit higher payout

¹¹ We also perform regressions for dependent variable ROA and Tobin's Q (independent variable CCC), and obtain similar results. In the described robustness checks, to save space, the results of the dependent variable ROA and Tobin's Q (independent variable CCC) are not tabulated.

ratios than constrained firms are. For each country and year, we classify firms into two groups based on the median firm size and dividend payout: small or low dividend payout firms (below the median) are classified as the financially constrained group, whereas large or high dividend payout firms (above the median) are classified as the financially unconstrained group. The results indicate that in both the financially constrained and financially unconstrained groups, *IndAdjCCC* ($IndAdjCCC \times LowCCC$) exhibits significantly negative (positive) relationships with two types of firm performance variables, thereby supporting that an aggressive liquidity management policy can enhance firm performance, and this effect reduces or reverses when firms exist at the lower CCC level.¹²

106.6 Conclusion

WCM is crucial for a firm's operating performance and corporate value. However, most existing literature on corporate finance has discussed issues regarding the relationship between long-term financial decisions, such as capital structure and capital expenditure, and corporate performance. Previous studies have seldom explored issues regarding liquidity management; hence, we conducted a global empirical analysis of the relationship between the cash conversion cycle (CCC) and corporate performance by adopting enterprises from different countries as samples.

Our sample consists of 46 countries, 31,612 companies, and 266,547 firm-year observations for the period of 1994–2011. The results indicate that industry-adjusted CCCs exhibit significantly negative relationships with industry-adjusted ROAs and industry-adjusted Tobin's Q, and that the negative relationships diminish or reverse when industry-adjusted CCC is below 0. This finding indicates that firms can shorten their CCC to increase profitability and value; however, this effect reduces or reverses when firms exist at the lower CCC level. Furthermore, the results remain unchanged after accounting for endogeneity and controlling for changes in macroeconomic environments, economic development status, financial crises, corporate governance, and financial constraints.

The CCC of WCM is a widely used metric for gauging the effectiveness of a firm's management and the intrinsic need for external financing (Wang 2019). Jalal and Khaksari (2020) suggest that short-term financial management is one of the three most important responsibilities of corporate financial managers, along with capital budgeting and long-term financing activities. Our study contributes to understanding the relationship between CCCs and firm performance, which can consequently help firms establish financial policies. Therefore, WCM is vital because of its impact on

¹² We also perform regressions for dependent variable ROA and Tobin's Q (independent variable CCC), and obtain similar results. In the described robustness checks, to save space, we do not tabulate the results of dependent variable ROA and Tobin's Q (independent variable CCC).

firm performance. The results also can help multinational companies to determine allocation proportions for short-term assets and capital.

Acknowledgments I gratefully acknowledge financial support from the National Science Council of Taiwan (NSC 100-2410-H-468-015).

References

- Afza, T., and M.S. Nazir. 2008. Working capital approaches and firm's returns in pakistan. *Pakistan Journal of Commerce and Social Sciences* 1: 25–36.
- Agrawal, A., and C.R. Knoeber. 1996. Firm performance and mechanisms to control agency problems between managers and shareholders. *Journal of Financial and Quantitative Analysis* 31: 377–397.
- Aktas, N., E. Croci, and D. Petmezas. 2015. Is working capital management value-enhancing? Evidence from firm performance and investments. *Journal of Corporate Finance* 30: 98–113.
- Almeida, H., M. Campello, and M.S. Weisbach. 2011. Corporate financial and investment policies when future financing is not frictionless. *Journal of Corporate Finance* 17: 675–693.
- Ang, J., and A. Smedema. 2011. Financial flexibility: Do firms prepare for recession. *Journal of Corporate Finance* 17: 774–787.
- Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.
- Baños-Caballero, S., P.J. García-Teruel, and P. Martínez-Solano. 2010. Working capital management in SMEs. *Accounting and Finance* 50: 511–527.
- . 2012. How does working capital management affect the profitability of Spanish SMEs? *Small Business Economics* 39: 517–529.
- Bates, T.W., K.M. Kahle, and R.M. Stulz. 2009. Why do U.S. firms hold so much more cash than they used to? *Journal of Finance* 64: 1985–2021.
- Blinder, A.S., and L.J. Maccini. 1991. The resurgence of inventory research: What have we learned? *Journal of Economic Surveys* 5: 291–328.
- Boisjoly, R.P., T.E. Conine Jr., and M.B. McDonald IV. 2020. Working capital management: Financial and valuation impacts. *Journal of Business Research* 108: 1–8.
- Booth, L., V. Aivazian, A. Demircuc-Kunt, and V. Maksimovic. 2001. Capital structures in developing countries. *Journal of Finance* 56: 87–130.
- Campello, M., J.R. Graham, and G.R. Harvey. 2010. The real effects of financial constraints: Evidence from a financial crisis. *Journal of Financial Economics* 97: 470–487.
- Carpenter, R.E., S.M. Fazzari, and B.C. Petersen. 1994. Inventory investment, internal-finance fluctuations, and business cycle. *Brookings Papers on Economic Activity* 25: 75–135.
- Céspedes, J., M. González, and C.A. Molina. 2010. Ownership and capital structure in Latin America. *Journal of Business Research* 63: 248–254.
- Chen, K.C.W., Z. Chen, and K.C.J. Wei. 2009. Legal protection of investors, corporate governance, and the cost of equity capital. *Journal of Corporate Finance* 15: 273–289.
- Chiou, J.R., L. Cheng, and H.W. Wu. 2006. The determinants of working capital management. *Journal of American Academy of Business* 10: 149–155.
- Cho, M.H. 1998. Ownership structure, investment, and the corporate value: An empirical analysis. *Journal of Financial Economics* 47: 103–121.
- Core, J.E., R.W. Holthausen, and D.F. Larcker. 1999. Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics* 51: 371–406.
- Core, J.E., W.R. Guay, and T. Rusticus. 2006. Does weak governance cause weak stock returns? An examination of firm operating performance and investors' expectations. *Journal of Finance* 61: 655–687.
- Cuñat, V. 2007. Trade credit: Suppliers as debt collectors and insurance providers. *Review of Financial Studies* 20: 491–527.

- Czyzewski, A.B., and D.W. Hicks. 1992. Hold onto your cash. *Management Accounting* 73: 27–30.
- De Vito, A., and J.P. Gómez. 2020. Estimating the COVID-19 cash crunch: Global evidence and policy. *Journal of Accounting and Public Policy* 39 (2): 106741.
- Deloof, M. 2003. Does working capital management affect profitability of Belgian firms? *Journal of Business Finance and Accounting* 30: 573–587.
- Djankov, S., R. La Porta, F. Lopez De-Silanes, and A. Shleifer. 2008. The law and economics of self-dealing. *Journal of Financial Economics* 88: 430–465.
- Doong, S.C., H.G. Fung, and J.Y. Wu. 2011. Are social, financial, and human capital value enhancing? Evidence from Taiwanese firms. *International Review of Economics and Finance* 20: 395–405.
- Emery, G. 1987. An optimal financial response to variable demand. *Journal of Financial and Quantitative Analysis* 22: 209–225.
- Erickson, T., C.H. Jiang, and T.M. Whited. 2014. Minimum distance estimation of the errors-in-variables model using linear cumulant equations. *Journal of Econometrics* 183: 211–221.
- Fama, E.F., and K.R. French. 1997. Industry costs of equity. *Journal of Financial Economics* 43: 153–193.
- Fazzari, S.M., and B.C. Petersen. 1993. Working capital and fixed investment: New evidence on financing constraints. *RAND Journal of Economics* 24: 328–342.
- Fazzari, S., R.G. Hubbard, and B.C. Petersen. 1988. Financing constraints and corporate investment. *Brookings Papers on Economic Activity* 1: 141–195.
- Filbeck, G., and T.M. Krueger. 2005. An analysis of working capital management results across industries. *American Journal of Business* 20: 11–18.
- García-Teruel, P.J., and P. Martínez-Solano. 2007. Effects of working capital management on SME profitability. *International Journal of Managerial Finance* 3: 164–177.
- González, V.M. 2013. Leverage and corporate performance: International evidence. *International Review of Economics and Finance* 25: 169–184.
- Hager, H.C. 1976. Cash management and the cash cycle. *Management Accounting* 57: 19–21.
- Hail, L., and C. Leuz. 2006. International differences in the cost of equity capital: Do legal institutions and securities regulation matter? *Journal of Accounting Research* 44: 485–531.
- He, Z., and W. Xiong. 2012. Rollover risk and credit risk. *Journal of Finance* 67: 391–430.
- Hill, M.D., G.W. Kelly, and G.W. Highfield. 2010. Net operating working capital behavior: A first look. *Financial Management* 39: 783–805.
- Howorth, C., and P. Westhead. 2003. The focus of working capital management in UK small firms. *Management Accounting Research* 14: 94–111.
- Jalal, A., and S. Khaksari. 2020. Cash cycle: A cross-country analysis. *Financial Management* 49: 635–671.
- Jose, M.L., C. Lancaster, and J.L. Stevens. 1996. Corporate returns and cash conversion cycles. *Journal of Economics and Finance* 20: 33–46.
- Kamath, R. 1989. How useful are common liquidity measures? *Journal of Cash Management* 9: 24–28.
- Kashyap, A.K., O.A. Lamont, and J.C. Stein. 1994. Credit conditions and the cyclical behavior of inventories. *Quarterly Journal of Economics* 109: 565–592.
- Kieschnich, R., M. LaPlante, and R. Moussawi. 2006. *Corporate working capital management: Determinants and consequences*. Working paper.
- Kim, Y. 2005. Board network characteristics and firm performance in Korea. *Corporate Governance: An International Review* 13: 800–808.
- Klapper, L. 2006. The role of factoring for financing small and medium enterprises. *Journal of Banking and Finance* 30: 3111–3130.
- Kroes, J.R., and A.S. Manikas. 2014. Cash flow management and manufacturing firm financial performance: A longitudinal perspective. *International Journal of Production Economics* 148: 37–50.

- Kuan, T.H., C.S. Li, and C.C. Liu. 2012. Corporate governance and cash holdings: A quantile regression approach. *International Review of Economics and Finance* 24: 303–314.
- Kusnadi, Y., and K.C.J. Wei. 2011. The determinants of corporate cash management policy: Evidence from around the world. *Journal of Corporate Finance* 17: 725–740.
- Lee, S.Y. 2015. The relationship between working capital management and profitability: Evidence from Korean shipping industry. *Journal of Navigation and Port Research* 39: 261–266.
- Lie, E. 2005. Operating performance following dividend decreases and omissions. *Journal of Corporate Finance* 12: 27–53.
- Lin, R.Y., and M.X. Fu. 2017. Does institutional ownership influence firm performance? Evidence from China. *International Review of Economics and Finance* 49: 17–57.
- Lin, Q., and X. Lin. 2021. Cash conversion cycle and aggregate stock returns. *Journal of Financial Markets* 52: 100560.
- Lin, J.R., C.J. Wang, D.W. Chou, and F.C. Chueh. 2013. Financial constraint and the choice between leasing and debt. *International Review of Economics and Finance* 27: 171–182.
- Lskavyan, V., and M. Spatareanu. 2006. Ownership concentration, market monitoring and performance: Evidence from the UK, the Czech Republic and Poland. *Journal of Applied Economics* 9: 91–104.
- Luo, M. 2011. A bright side of financial constraints in cash management. *Journal of Corporate Finance* 17: 1430–1444.
- Manikas, A.S., and P.C. Patel. 2016. Managing sales surprise: The role of operational slack and volume flexibility. *International Journal of Production Economics* 179: 101–116.
- McConnell, J.J., and C.J. Muscarella. 1985. Corporate capital expenditure and the market value of the firm. *Journal of Financial Economics* 14: 399–422.
- Michaelas, N., F. Chittenden, and P. Poutziouris. 1999. Financial policy and capital structure choice in UK SMEs: Evidence from company panel data. *Small Business Economics* 12: 113–130.
- Molina, C., and L. Preve. 2009. Trade receivables policy of distressed firms and its effect on the cost of financial distress. *Financial Management* 38: 663–686.
- Morck, R., A. Shleifer, and R.W. Vishny. 1988. Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics* 20: 293–315.
- Myers, S., and N. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13: 187–221.
- Newey, W.K., and K.D. West. 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55: 703–708.
- Ng, C.K., J.K. Smith, and R.L. Smith. 1999. Evidence on the determinants of credit terms used in interfirm trade. *Journal of Finance* 54: 1109–1129.
- Opler, T., L. Pinkowitz, R. Stulz, and R. Williamson. 1999. The determinants and implications of corporate cash holdings. *Journal of Financial Economics* 52: 3–46.
- Oseifuah, E.K., and A. Gyekye. 2016. Cash conversion cycle theory and corporate profitability: Evidence from non-financial firms listed on the Johannesburg stock exchange. *Journal of Accounting and Management* 6: 3.
- Padachi, K. 2006. Trends in working capital management and its impact on firms' performance: an analysis of Mauritian small manufacturing firms. *International Review of Business Research Papers* 2: 45–58.
- Peters, R.H., and L.A. Taylor. 2017. Intangible capital and the investment-q relation. *Journal of Financial Economics* 123 (2): 251–272.
- Petersen, M., and R. Rajan. 1997. Trade credit: Theories and evidence. *Review of Financial Studies* 10: 661–691.
- Pombo, C., and R. Taborda. 2017. Stock liquidity and second blockholder as drivers of corporate value: Evidence from Latin America. *International Review of Economics and Finance* 51: 214–234.
- Raheman, A., and M. Nasr. 2007. Working capital management and profitability – case of Pakistani firms. *International Review of Business Research Papers* 3: 275–296.

- Reinhart, C.M., and K.S. Rogoff. 2011. From financial crash to debt crisis. *American Economic Review* 101: 1676–1706.
- Riddiough, T.J., and Z. Wu. 2009. Financial constraints, liquidity management and investment. *Real Estate Economics* 37: 447–481.
- Sensini, L., and M. Vazquez. 2021. Effects of working capital management on SME profitability: Evidence from an emergent economy. *International Journal of Business and Management* 16 (4): 85–95.
- Shin, H.H., and L. Soenen. 1998. Efficiency of working capital management and corporate profitability. *Financial Practice and Education* 8: 37–45.
- Shleifer, A., and D. Wolfenzon. 2002. Investor protection and equity markets. *Journal of Financial Economics* 66: 3–27.
- Smith, K. 1980. Profitability versus liquidity tradeoffs in working capital management. In *Readings on the management of working capital pp*, ed. K.V. Smith, 549–562. St Paul: West Publishing Company.
- Smith, J.K. 1987. Trade credit and informational asymmetry. *Journal of Finance* 42: 863–872.
- Soenen, L.A. 1993. Cash conversion cycle and corporate profitability. *Journal of Cash Management* 13: 53–57.
- Tong, H., and S.J. Wei. 2011. The composition matters: Capital inflows and liquidity crunch during a global economic crisis. *Review of Financial Studies* 24: 2023–2052.
- Tsai, C.L. 2014. The effects of monetary policy on stock returns: Financing constraints and “informative” and “uninformative” FOMC statements. *International Review of Economics and Finance* 29: 273–290.
- Uyar, A. 2009. The relationship of cash conversion cycle with firm size and profitability: An empirical investigation in Turkey. *International Research Journal of Finance and Economics* 24: 186–193.
- Wang, Y. 2002. Liquidity management, operating performance, and corporate value: Evidence from Japan and Taiwan. *Journal of Multinational Financial Management* 12: 159–169.
- Wang, B. 2019. The cash conversion cycle spread. *Journal of Financial Economics* 133: 472–497.
- Whited, T.M. 1992. Debt, liquidity constraints and corporate Investment: Evidence from panel data. *Journal of Finance* 47: 1425–1460.
- Wilner, B.S. 2000. The exploitation of relationships in financial distress: The case of trade credit. *Journal of Finance* 55: 153–178.
- Zimon, G., and H. Tarighi. 2021. Effects of the COVID-19 global crisis on the working capital management policy: Evidence from Poland. *Journal of Risk and Financial Management* 14 (4): 169.



How Consistent Are the Judges of Portfolio Performance?

107

Matthew Brigida, Chin W. Yang, and Ken Hung

Abstract

This analysis tests whether the rank ordering of a set of portfolios, with varying number of assets, differs depending on the performance measure used. To test this hypothesis, we construct a Friedman test over 4950 portfolios (50 portfolios for each portfolios sizes 2–100) of S&P 500 stocks. We find that while measures may differ in their sensitivity to portfolio size, all measures consistently prefer larger portfolios. Therefore, the rank ordering of a set of portfolios of various sizes is consistent across performance measures.

Keywords

Portfolio theory · Portfolio performance evaluation

JEL Classification

G10 · G11

M. Brigida (✉)
SUNY Polytechnic Institute, Utica, NY, USA
e-mail: matthew.brigida@sunyit.edu

C. W. Yang
Clarion University of Pennsylvania, Clarion, PA, USA
National Chung Cheng University, Chia-yi, Taiwan
e-mail: yang@clarion.edu

K. Hung
Division of International Banking and Finance Studies, Texas A&M International University,
Laredo, TX, USA
e-mail: ken.hung@tamiu.edu

107.1 Introduction

The goal of this analysis is to compare the rankings of four popular portfolio performance measures as portfolio size increases. These measures considered are the Sharpe ratio (Sharpe [1966]), information ratio (Goodwin [1988]), Treynor ratio (Treynor [1965]), and Jensen's alpha (Jensen [1969]). Note, these are four unique measures – they are not linear transformations of each other.

This analysis is intended to assist practitioners (portfolio managers and other investment professionals) when ranking portfolios of differing sizes. Specifically, it will give them a better understanding of how a varying number of assets in a set of potential portfolios will affect the rank ordering of the set provided by each performance measure. Thus, given a set of portfolios of differing sizes, this analysis will answer the question: Is the rank ordering of this set of portfolios dependent on the performance measure used?

Note our focus on ranks and not the value of the performance measure. First, values of one performance measure cannot be compared to another. Moreover, for practical purposes, it is the ranking of the portfolios which matters, and not the comparative values of the performance measures. So to compare the behavior of various measures, we compare the respective rankings they provide, and test whether these rankings are statistically different among performance measures.

Previous studies have shown portfolio managers may trade in such way as to increase the value of performance measures, despite not having any private information. For example, Ferson and Siegel (2001) and Lhabitabt (2000) use separate methods to increase the value of a portfolio's Sharpe ratio. We do not consider such trading strategies, since the portfolio manager is using the measures as a way to rank portfolios for potential investment.

107.2 Construction

Our dataset consists of daily log returns for every stock in the S&P 500 index, as well as returns for major indices themselves (S&P 500, Wilshire 5000, Dow Jones Industrial Average, etc.) over the 3 years beginning January 2010 and ending December 2012. The log returns were calculated using stock price data from interactive brokers.

First, we created 50 random portfolios of size two from the set of stocks in the S&P 500. We then calculate each performance measure for each of the 50 random portfolios. We repeat this procedure for portfolio sizes up to 100. This gives us the performance measure values for each of the 50 random portfolios for each portfolio size 2–100 (4950 portfolios).

We then rank the 4950 portfolios using each measure. So for each performance measure we have a ranking of the random portfolios from 1 to 4950, where 1 denoted the portfolio with the highest ranking (most preferred). Letting $r_{n,i}$ denote the rank of the i^{th} portfolio of size n (where $2 \leq n \leq 50$). Then $r_{n,i}$ is an integer between 1 and 4950 inclusive.

Lastly for each portfolio size we sum the ranks of the portfolios of that size (for each n denote this as $\varphi_n = \sum_{i=1}^{50} r_{n,i}$). So for all 99 portfolio sizes we have a sum of the rankings of the 50 random portfolios of that size (φ_n) for each portfolio measure.

This methodology allows us to compare how each portfolio measure ranking changes as the portfolio size changes. A plot of the sum of each ranking (φ_n) for each measure, over the 99 portfolio sizes, is in Fig. 1 below. Figure 1 also contains a linear trendline showing how each portfolio measure is affected by increasing portfolio size.

The figure shows that the Sharpe and information ratios are more sensitive to portfolio size than the Treynor ratio and Jensen’s alpha. We see that, on average, φ_n is 493 and 330 less than φ_n for the Sharpe and information ratios, respectively. This compares with 27 and 15 for the Treynor ratio and Jensen’s alpha, respectively.

Note that a decrease in φ_n means an increase in rank (because for each portfolio measure the highest value has rank 1 and lowest value rank 4950). Thus, our results so far are evidence that Sharpe and information ratios have a tendency to give a higher rank to larger portfolios. The Treynor ratio and Jensen’s alpha have the same tendency, though to a much less degree.

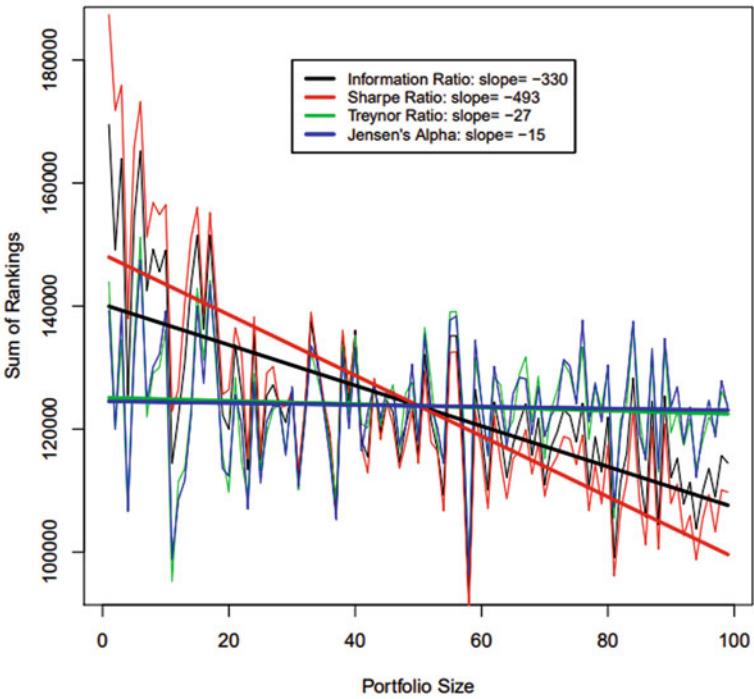


Fig. 1 Sum of portfolio rankings over portfolio sizes 1–100 with trendlines. The negative slope of each trendline shows that each portfolio measure tends to rank larger portfolios higher

107.3 The Friedman Test

In this section we will formally test whether the portfolio measures form consistent rankings of the portfolios. We may think of each portfolio measure as a judge, and each portfolio as a contestant. We wish to test whether each judge will rank the contestants consistently across portfolio sizes. Or rather, we are testing whether a particular judge has a bias toward large, or small, portfolios.

To test this we will use the Friedman test (Friedman 1937, 1939, 1940) which is designed to test whether judges' ranking are consistent with one another. A classic example is to test whether n judges rank k wines consistently. In our case the n judges are our portfolio measures, and the k wines are our 4950 portfolios. Note, the Friedman test is nonparametric, which is necessary given nonnormality of some performance measures.

To conduct the Friedman test we sum the sum of rankings over each portfolio measure, for each portfolio size. For example, consider the 3 stock portfolio size in Table 1 below. We sum each column across the performance measures, and have a sum of $151,887 + 135,609 + 116,304 + 105,176 = 508,976$. For each portfolio size we denote this value as ϖ_n .

The null hypothesis is that the average ranking is equal across all portfolio sizes, that is, $\varpi_2 = \varpi_3 = \dots = \varpi_{100}$. Under this null hypothesis our test statistic $Q = \frac{12S}{nk\lambda^2(n\lambda+1)} X_{n-1}^2$, where n is the number of portfolios (here 99), k is the number of performance measures (here 4), λ is the number of repetitions (here 50), and $S = \sum_{w_i=2}^{100} \varpi_i^2 - \frac{k^n \lambda^2 (n\lambda+1)^2}{4}$.

If we reject the null of the Friedman test, this is evidence that the judges consistently rank the contestants. Conversely if we cannot reject the null this is evidence that the rankings are inconsistent.

107.3.1 Results

Using the Friedman test, from portfolio size 2–100, we reject the null at the less than 0.1% level of significance (the test statistic $Q = 462.67$). This is evidence that all judges are on average ranking the portfolios consistently, that is, all judges consistently prefer larger portfolios. Further, the null hypothesis is also rejected if the test is run over any set of 10 consecutive portfolio sizes.

Table 1 Structure of Friedman test

Measure/portfolio size	2	3	4	...	100
Sharpe ratio	171,463	151,887	142,251	...	104,580
Information ratio	147,382	135,609	131,635	...	108,492
Treynor ratio	118,937	116,304	129,489	...	116,066
Jensen's alpha	114,552	105,176	122,702	...	119,816
ϖ_n	552,334	508,976	526,077	...	448,954

107.4 Conclusion

Our four portfolio measures are consistent in their preference for larger portfolios, i.e., the measures are affected by portfolio size, but in a consistent manner. They only differ in their sensitivities to portfolio size. That is, the portfolio rankings provided by the Sharpe and information ratios afford a greater preference for larger portfolios than do the Treynor ratio and Jensen's alpha.

In sum, however, the answer to our question posed above is: On average, the rank ordering of a given set of portfolios of differing numbers of assets is not dependent on the performance measure used. That is, while these are four unique measures, each of the performance measures is in itself adequate.

References

- Ferson, W., and A. Siegel. 2001. The efficient use of conditioning information in portfolios. *Journal of Finance* 3: 967–982.
- Friedman, Milton. December 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association (American Statistical Association)* 32 (200): 675–701.
- . March 1939. A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association (American Statistical Association)* 34 (205): 109.
- . March 1940. A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics* 11 (1): 86–92.
- Goodwin, T. (1988). The information ratio. *Financial Analysts Journal* July/August, 34–43.
- Jensen, M. 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* 42: 167–247.
- Lhabitant, F. 2000. Derivatives in portfolio management: Why beating the market is easy. *Derivatives Quarterly* 6: 39–45.
- Sharpe, W.F. 1966. Mutual fund performance. *Journal of Business* 39: 119–138.
- Treynor, J. 1965. How to rate management of investment funds. *Harvard Business Review* 43: 63–75.

Risk Aversion and the Value of Information for Investors

108

Richard E. Kihlstrom

Abstract

The chapter provides a selective survey of the literature relating risk aversion to the value of information. The focus is on several papers, Cabrales et al. (*American Economic Review* 103:360–77, 2013; *Journal of Economic Theory* 170:266–288, 2017) and Losq and Sobti (Demand for information and risk aversion: Some results in a portfolio choice context, unpublished working paper, McGill University, 1985) in which more risk averse investors are shown to put a lower value on information. The intuition for this result suggested by the CARA case considered by Losq and Sobti (Demand for information and risk aversion: Some results in a portfolio choice context, unpublished working paper, McGill University, 1985) is the more conservative investment strategies chosen by highly risk averse investors. The argument used in Losq and Sobti (Demand for information and risk aversion: Some results in a portfolio choice context, unpublished working paper, McGill University, 1985) is quite specific to the CARA case. In Cabrales et al. (*American Economic Review* 103:360–77, 2013; *Journal of Economic Theory* 170:266–288, 2017), the result is a consequence of the assumptions that imply uninformed investors bear no risk. When the investor bears no risk, information acquisition is a risky investment. These results are related to a gambling example provided by a 2001 literature survey in Gollier (*The economics of risk and time*. MIT Press, 2001). The chapter concludes with several CRRA examples that suggest the CARA results can be extended. There is also a digression on the point made in Cabrales et al. (*American Economic Review* 103:360–77, 2013) on the role of entropy in this literature.

Keywords

Information value · Risk aversion · Portfolio theory

R. E. Kihlstrom (✉)

The Wharton School of the University of Pennsylvania, Philadelphia, PA, USA

e-mail: kihlstro@wharton.upenn.edu

JEL ClassificationC44 · D81 · D83 · G11 · G14

108.1 Introduction

Are more risk averse investors willing to pay more for information? Intuition would seem to suggest they are. Two recent papers by Cabrales, Gossner, and Serrano in the *American Economic Review* (Cabrales et al. 2013) and *Journal of Economic Theory* (Cabrales et al. 2017) consider this question and arrive at the opposite result, although under quite special circumstances. These papers are among the latest contributions to the literature that studies the effect of risk aversion on the value of information. This literature has a long but sporadic history that began soon after the introduction of the Arrow-Pratt measures of absolute and relative risk aversion in Arrow (1965), and Pratt (1964). In Chap. 24 of Gollier (2001), Gollier provided a discussion of the state of this literature as of 2001. This chapter is a very selective updated survey that is partially motivated by the Cabrales et al. contributions.

The *American Economic Review* paper by Cabrales, Gossner, and Serrano focused on the place of entropy in this literature. The subsequent paper by the same authors extended the results of the *American Economic Review* paper. The crucial feature of the Cabrales et al. setting is that all risk averse, uninformed investors choose not to take risk. The Cabrales et al. result was partially anticipated by Gollier who, in his 2001 survey, presented an example of a risk averse gambler who also values information less when he is more risk averse. Gollier also obtained this result for the case in which the gambler was so risk averse that he chooses not to gamble when he is uninformed. Gollier went on to note, however, that if the gambler is not very risk averse, he will gamble when he is uninformed and, in such a case, he values information more when he becomes somewhat more risk averse, in contrast to the Cabrales et al. result.

A second motivation for this chapter is to describe some results that had been obtained in an unpublished and, therefore, underappreciated paper written by Losq and Sobti (1985). That paper considered the portfolio problem faced by an investor who maximizes the expected value of a utility function that exhibits constant absolute risk aversion (CARA). In that setting Losq and Sobti also demonstrated that, under general conditions, more risk averse investors value information less. This is a result that, as Gollier's survey discusses, had been obtained after 1985 by other papers. The other papers that obtained this result were, however, only able to do so by also assuming that the distribution of returns was normal, an assumption not required by Losq and Sobti.

108.1.1 Outline of the Chapter

We begin with a brief discussion of the Cabrales et al. main result. We digress to discuss the economic conditions Cabrales et al. describe under which the expression

for the value of information is a ratio of certainty equivalents computed using entropy. They note, in particular, that entropy arises in the expression for the investor's maximum expected utility when the investor's utility function has a CRRA of one. In a 1962 *Annals of Mathematical Statistics* paper, DeGroot (1962) treated (negative) entropy as an example of an "uncertainty function" and obtained generalized versions of the information provided by an experiment using uncertainty functions. DeGroot's paper was a follow-up to a 1956 *Annals of Mathematical Statistics* paper (Lindley 1956) in which Lindley had used entropy to describe the information provided by an experiment. We are motivated by DeGroot's interpretation to derive the "uncertainty functions" other than entropy that arise as the investor's maximum expected utility when the investor's utility function is in the CRRA class, but with a level of relative risk aversion other than one. We then use these uncertainty functions to compute the value of information in the economic setting considered by Cabrales et al.

The discussion then moves on to the Losq-Sobti result for the case of CARA utility functions. When Arrow and Pratt introduced their measures of risk aversion, they noted that an expected utility maximizing investor who simply chooses between a safe and risky asset invests less in the risky asset if his risk aversion measure increases. Losq and Sobti note the strong sense in which this is true for investors whose utility functions are CARA. They describe an explicit relationship between the absolute risk aversion measure and the investor's risky asset position. This relationship forms the basis for their demonstration that an increase in risk aversion reduces the value of information. So Losq and Sobti's analysis suggests that the more cautious investment strategy followed by the more risk averse CARA investors results in information being less important and, hence, less valuable for these investors. This interpretation is consistent with Gollier's example in which a gambler is unable to "gamble more or less cautiously." He either gambles or he does not. In that example, in the range of risk aversion levels for which the uninformed choice is to gamble, the value of information about the gambling outcome increases with risk aversion.

In an attempt to begin an investigation of the extent to which the intuition gained from the CARA case can be extended, we conclude the chapter by considering some examples in which the utility function is in the CRRA class. In contrast to the CARA case, there is no explicit relationship for the CRRA class between the relative risk aversion measure and the investor's risky asset position. We are, thus, motivated to limit attention to numerical examples. The numerical results obtained for the examples prove to be consistent with the intuition developed from the CARA case and Gollier's gambling example. Specifically, we note that, when the CRRA investors are not faced with borrowing or lending constraints that prevent them from behaving more aggressively when they are less risk averse, they value information more when they are less risk averse. When borrowing and/or lending constraints are imposed we do find ranges of the relative risk aversion measure over which less risk averse investors value information less. But on those ranges, the borrowing and/or lending constraints do prevent more aggressive behavior by less risk averse investors.

108.1.2 Some Clarifying Remarks

First, we should note that, throughout the discussion, we assume that investors maximize expected utility. The risk aversion measures considered are the usual Arrow-Pratt absolute and relative measures for utility functions of a single variable, final wealth. The discussion takes place in a static, partial equilibrium setting. This means that asset prices do not change when the investor becomes informed; we are not discussing “rational expectations.” We will make further comments on this point as the discussion proceeds.

108.2 A General Framework

This section describes the general framework that we use to introduce the definition of the value of information used in this chapter. This general framework is given different interpretations as we proceed with our discussion.

We treat the investor’s problem as a general decision problem in which the investor chooses a portfolio $c = \hat{c}(u, p, W, \mu)$ where

$$\hat{c}(u, p, W, \mu) = \arg \max_{c \in C(p, W)} \sum_{x \in X} u(I(x, c)) \mu(x)$$

and where x is the state of the world that determines the return to an investment in the risky assets. The state of the world, x , is the realization of a random variable, $\tilde{x}$, that (for ease of exposition and without loss of generality) takes values in a finite set X . The probability of $\tilde{x} = x$ is denoted by $\mu(x)$, and

$$\Pr(\tilde{x} \in X) = 1.$$

We use $I(x, c)$ to denote the investor’s final wealth when he invests in the portfolio c and the state of the world is x . The investor’s utility function of I is $u(I)$ which is assumed to be strictly concave. The investor’s initial wealth is W and p denotes the vector of asset prices. The set of possible portfolios is the investor’s wealth constraint, $C(p, W)$.

Remark 2 *Because, for each portfolio choice, c , expected utility is a linear function of μ , the function*

$$\varphi(u, p, W, \mu) = \max_{c \in C(p, W)} \sum_{x \in X} u(I(x, c)) \mu(x)$$

is a convex function of μ .

We now introduce the possibility that the investor can observe a signal $\tilde{s}$ that is correlated with $\tilde{x}$ and can make use of the information about $\tilde{x}$ provided by this

observation to make a better portfolio choice. The sample space of $\tilde{s}$ is denoted by S . We denote the unconditional distribution of $\tilde{x}$ by μ_0 . For each $s \in S$, let μ_s be the conditional distribution of $\tilde{x}$ given that $\tilde{s} = s$. When he has observed $\tilde{s} = s$, the investor chooses,

$$\hat{c}(u, p, W, \mu_s) = \arg \max_{c \in C(p, W)} \sum_{x \in X} u(I(x, c)) \mu_s(x).$$

Conditional on having observed that $\tilde{s} = s$, the investor's expected utility is

$$\varphi(u, p, W, \mu_s) = \arg \max_{c \in C(p, W)} \sum_{x \in X} u(I(x, c)) \mu_s(x).$$

It must be emphasized that, even after the investor observes that $\tilde{s} = s$, the prices remain unchanged; they are the same as they were when the investor was uninformed. This is what we meant earlier when we noted that we were discussing a "partial equilibrium setting" and not considering "rational expectations." In the rational expectations literature prices do change when investors become informed.

Once he knows that he will observe $\tilde{s}$ before choosing c , but has not yet observed it, the *ex ante* expected utility of information is

$$E\varphi(u, p, W, \mu_s),$$

where, of course, the unconditional distribution of $\tilde{s}$ is used to compute the expectation. Because $\varphi(u, p, W, \mu)$ is a convex function of μ , Jensen's inequality implies that

$$E\varphi(u, p, W, \mu_s) \geq \varphi(u, p, W, E\mu_s). \quad (1)$$

Since

$$E\mu_s = \mu_0,$$

(1) asserts that

$$E\varphi(u, p, W, \mu_s) \geq \varphi(u, p, W, \mu_0).$$

This means that the expected value of making an informed decision exceeds the expected utility of making the decision without information.

The argument just presented that the expected utility of making an informed decision exceeds the expected utility of making an uninformed decision is rather old. It appeared in an unpublished RAND Research Memorandum (Bonnenblust and Sherman 1949) in 1949. Blackwell used it in his paper Blackwell (1953) on the comparison of experiments.

We can now define the value of information, $V = V(u, p, W, \mu_0, \tilde{s})$, as

$$E\varphi(u, p, W - V, \mu_{\tilde{s}}) = \varphi(u, p, W, \mu_0).$$

It will also be useful to define $v = v(u, p, W, \mu_0, \tilde{s}) = \frac{V(u, p, W, \mu_0, \tilde{s})}{W}$, which is simply the value of information as a fraction of wealth.

108.3 An Exposition of the Cabrales et al. Contribution

Cabrales et al. (2013, 2017) showed that, for a very special case,

$$V(u^1, p, W, \mu_0, \tilde{s}) < V(u^2, p, W, \mu_0, \tilde{s})$$

if the utility function u^1 is more risk averse than the utility function u^2 . In Cabrales et al. (2013) they actually discussed and considered the case in which

$$u^2(I) = \log(I).$$

But the result is more general and the same authors generalized it in Cabrales et al. (2017).

108.3.1 The Setting

In Cabrales et al. (2013, 2017) markets are assumed to be complete and an additional assumption was made about asset prices. This additional assumption implies that any risk averse uninformed investor takes no risk. We exposit their results by assuming a market for Arrow-Debreu contingent claims. A contingent claim that pays \$1 in state x and nothing in any other state has a price p_x . In this case, the investor's budget set is

$$C(p, W) = \left\{ c \in R_n : \sum_{x \in X} p_x c_x = W \right\}$$

The assumption that implies any risk averse uninformed investor takes no risk is that, for all x , $p_x = \mu_0(x)$. They referred to this as the assumption of "risk-neutral pricing," but this assumption is stronger than what is usually referred to as "risk-neutral pricing." It might more accurately be referred to as "actuarially fair pricing." It effectively means that asset prices are determined in a market in which the marginal investor is risk neutral and for whom the probability distribution of the states is μ_0 . When the value of information is computed, it is assumed that the prices remain $p = \mu_0(\cdot)$ after the investor becomes informed. It is quite easy to see that, for any strictly concave utility function u ,

$$\widehat{c}(u, p = \mu_o, W, \mu_o) = W$$

and

$$\varphi(u, p = \mu_0, W, \mu_0) = u(W).$$

Thus, an investor who maximizes the expected value of u chooses to hold only the riskless asset and bears no risk.

108.3.2 Risk Aversion and Information Value in This Setting

Cabrales et al. show that

$$V(u^1, p = \mu_0, W, \mu_0, \widetilde{s}) < V(u^2, p = \mu_0, W, \mu_0, \widetilde{s})$$

if u^1 is more risk averse than u^2 . The acquisition of information in the Cabrales et al. setting is analogous to the acquisition of a valuable but risky investment opportunity. The proof of their result is the same as the proof of the well-known result that a more risk averse investor will put a lower value on such an opportunity.

Although the uninformed investor chooses not to bear risk when $p = \mu_0$, he does choose to bear risk when he becomes informed and his beliefs are updated so that $\mu_s \neq \mu_0$, but the prices are unchanged so that $p = \mu_0$. This is what makes information acquisition effectively the same as an investment in a profitable risky investment. The more risk averse investor is more averse to the risks he chooses to take on learning that $\mu_s \neq \mu_0$. These risks are compounded by the fact that, *ex ante*, the investor faces the risks of fluctuations in the conditional distribution μ_s .

108.3.3 Entropy and the Value of Information

In the discussion to follow we will not impose the assumption that

$$p = \mu_0(\cdot).$$

So we do not impose restrictions on the vector of contingent claims prices. When the utility function is

$$u(I) = \log(I),$$

the optimal portfolio of contingent claims is $\widehat{c}$, where

$$\widehat{c}_x(u, p, W, \mu) = \frac{\mu(x)W}{p_x}.$$

Thus,

$$\begin{aligned} \varphi(u, p, W, \mu) \\ = \sum_{x \in X} \log(\mu(x))\mu(x) - \sum_{x \in X} \log(p_x)\mu(x) + \log(W) \end{aligned}$$

It will be noted that the first term in this expression is the negative of entropy, a concave function of μ .

In this case,

$$\begin{aligned} E\varphi(u, p, W, \mu_s^-) &= E \left[\sum_{x \in X} \log(\mu_s^-(x))\mu_s^-(x) \right] \\ &\quad - \sum_{x \in X} \log(p_x)E\mu_s^-(x) + \log(W) \end{aligned}$$

where

$$\sum_{x \in X} \log(p_x)E\mu_s^-(x) = \sum_{x \in X} \log(p_x)\mu_0(x)$$

The value of information as a fraction of wealth, $v = \frac{V}{W}$, is the solution of

$$E\varphi(u, p, W(1-v), \mu_s^-) = \varphi(u, p, W, \mu_0)$$

which simplifies to

$$\begin{aligned} &\log(1-v) \\ &= \sum_{x \in X} \log(\mu_0)\mu_0(x) - E \left[\sum_{x \in X} \log(\mu_s^-)\mu_s^-(x) \right]. \end{aligned}$$

So, in the log utility case

$$v = 1 - \left(\frac{\exp \sum_{x \in X} \log(\mu_0)\mu_0(x)}{\exp E \left[\sum_{x \in X} \log(\mu_s^-)\mu_s^-(x) \right]} \right)$$

Note that the exponent in the denominator is simply the expected negative entropy of the possible posterior (conditional) distributions and the exponent in the numerator is the negative entropy of the prior (unconditional) distribution. Note also that v is independent of the price vector p and of wealth, W .

108.3.4 CRRA Utility and Information Value

In the preceding section we noted that, when

$$u(I) = \log(I),$$

and the relative risk aversion is, therefore, constant and equal to one, the function $\varphi(u, p, W(1 - v), \mu)$ is essentially the convex function that is the negative of entropy. We now derive the function φ for CRRA utility functions $u(c)$ for which the relative risk aversion is not one, that is, we now suppose that

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma},$$

where γ positive and unequal to one. We continue to analyze the value of information in the complete market setting described above.

In that case, the optimal demand for state x contingent claims is

$$\hat{c}_x(\gamma, p, W, \mu) = \frac{W \left[\frac{\mu(x)}{p_x} \right]^{\frac{1}{\gamma}}}{\sum_{x' \in X} \left[\frac{\mu(x')}{p_{x'}} \right]^{\frac{1}{\gamma}} p_{x'}}$$

and

$$\varphi(\gamma, p, W, \mu) = \frac{W^{1-\gamma}}{(1-\gamma)} \psi(\gamma, p, \mu)$$

where

$$\begin{aligned} \psi(\gamma, p, \mu) &= \left(\sum_{x \in X} \left[\frac{\mu(x)}{p_x} \right]^{\frac{1-\gamma}{\gamma}} \mu(x) \right)^{\gamma} \\ &= \left(\sum_{x \in X} p_x^{\frac{\gamma-1}{\gamma}} \mu(x)^{\frac{1}{\gamma}} \right)^{\gamma}. \end{aligned}$$

When $\gamma > 1$, φ equals a negative constant times the concave “uncertainty function” ψ . The value of information as a fraction of wealth solves

$$(1 - v)^{1-\gamma} E\psi(\gamma, p, \mu_s) = \psi(\gamma, p, \mu_0)$$

and

$$v(\gamma, p, \mu_0, \tilde{s}) = 1 - \frac{[\psi(\gamma, p, \mu_0)]^{\frac{1}{1-\gamma}}}{[E\psi(\gamma, p, \mu_s)]^{\frac{1}{1-\gamma}}}$$

where

$$\psi(\gamma, p, \mu) = \left(\sum_{x \in X} p_x^{\frac{\gamma-1}{\gamma}} \mu(x)^{\frac{1}{\gamma}} \right)^{\gamma}.$$

108.4 The CARA Case

While the CRRA utility functions have become the most widely used class of utility functions in the finance literature of recent years, the CARA class is commonly used in the theoretical market microstructure literature. As noted in the introduction, a 1985 unpublished paper (Losq and Sobti 1985) written by Losq and Sobti assumed constant absolute risk aversion and studied the impact of risk aversion on the value of information. They demonstrated that more risk averse investors facing a general portfolio problem value information less. While this is the same conclusion obtained by Cabrales et al., the logic of the argument used to obtain the result is quite different and the setting of the portfolio problem is more general. In this section, we exposit their result and attempt to clarify the underlying intuition.

108.4.1 The Simple Portfolio Problem

We assume that an investor, with absolute risk aversion α , can invest in a safe and risky asset and we now use c to denote the amount invested in the risky asset. We will assume that the investor cannot sell short or borrow to invest and we have $p = 1$. So the constraint set is

$$C(W) = \{0 \leq c \leq W\}.$$

The return on the safe asset is one and the excess return on the risky asset is x . We continue to assume that, x , is the realization of a random variable, $\tilde{x}$, that (for ease of exposition and without loss of generality) takes values in a finite set X . We continue to use $\mu(x)$ to denote the probability of $\tilde{x} = x$. No other distributional assumptions are imposed.

These assumptions imply that

$$\begin{aligned} \varphi(\alpha, W, \mu) &\equiv \varphi(u^\alpha, W, \mu) \\ &= \max_{0 \leq c \leq W} \sum_{x \in X} -\exp[-\alpha(W + cx)]\mu(x) \end{aligned}$$

The amount the investor invests in the risky asset is

$$\begin{aligned} \hat{c}(\alpha, W, \mu) &= \arg \max_{c \in [0, W]} -(\exp[-\alpha W]) \left[\sum_{x \in X} (\exp[-\alpha cx])\mu(x) \right] \\ &= \arg \max_{c \in [0, W]} -\sum_{x \in X} (\exp[-\alpha cx])\mu(x) \end{aligned}$$

So $\hat{c}(\alpha, W, \mu)$ is independent W and we can write $\hat{c}(\alpha, W, \mu)$ as $\hat{c}(\alpha, \mu)$. We should emphasize that this assumes that

$$\arg \max_{c \in [0, W]} - \sum_{x \in X} (\exp [-acx]) \mu(x) < W.$$

Of course, if this condition fails, $\widehat{c}(\alpha, W, \mu) = W$. So we are assuming an interior solution.

In fact, when we have an interior solution to the portfolio problem,

$$\widehat{c}(\alpha, \mu) = \frac{\widehat{c}(\alpha = 1, \mu)}{\alpha}$$

and the investor's maximum utility is simply

$$\begin{aligned} \varphi(\alpha, W, \mu) &= \max_{c \in [0, W]} - (\exp [-\alpha W]) \sum_{x \in X} (\exp [-acx]) \mu(x) \\ &= -(\exp [-\alpha W]) \psi(\mu) \end{aligned}$$

where

$$-\psi(\mu) = \max_{c \in [0, W]} - \sum_{x \in X} (\exp [-cx]) \mu(x).$$

The important feature of this expression is that the function ψ is independent of α !

At this point, we introduce the possibility that the investor can observe a signal $\tilde{s}$ that is correlated with $\tilde{x}$ and make use of the information about $\tilde{x}$ provided by his observation of $\tilde{s}$ to make a better decision. Again S is the sample space of $\tilde{s}$. For each $s \in S$, let μ_s be the conditional (posterior) distribution of $\tilde{x}$ given that $\tilde{s} = s$. When the investor has observed $\tilde{s} = s$, he invests

$$\widehat{c}(\alpha, \mu_s) = \frac{\widehat{c}(\alpha = 1, \mu_s)}{\alpha}$$

in the risky asset and his expected utility is

$$\psi(\alpha, W, \mu_s) = -(\exp [-\alpha W]) \psi(\mu_s)$$

Again we emphasize that this again assumes an interior solution.

The equation defining the value of information, $V = V(\alpha, W, \mu_0, \tilde{s})$,

$$E\varphi(\alpha, W - V, \mu_s) = \varphi(\alpha, W, \mu_0),$$

becomes

$$(\exp [-\alpha(W - V)]) E\psi(\mu_s) = (\exp [-\alpha W]) \psi(\mu_0).$$

Since

$$0 > -E\psi(\mu_s) \geq -\psi(\mu_0),$$

the value of information is

$$V(\alpha, W, \mu_0, \tilde{s}) = \frac{\log\left(\frac{\psi(\mu_0)}{E\psi(\mu_s)}\right)}{\alpha} \geq 0.$$

Note that $V(\alpha, W, \mu_0, \tilde{s})$ is decreasing in the absolute risk aversion measure α and is independent of wealth, W .

Also, the value of information as a fraction of wealth

$$v(\alpha, W, \mu_0, \tilde{s}) = \frac{V(\alpha, W, \mu_0, \tilde{s})}{W} = \frac{\log\left(\frac{\psi(\mu_0)}{E\psi(\mu_s)}\right)}{\alpha W}$$

is decreasing in wealth while the relative risk aversion, αI , is increasing. This follows from the observation that $V(\alpha, W, \mu_0, \tilde{s})$ is independent of W .

The fact that

$$\hat{c}(\alpha, \mu) = \frac{\hat{c}(\alpha = 1, \mu)}{\alpha}$$

means that a more risk averse investor simply holds a portfolio in which his investment in the risky asset is scaled back to account for his higher aversion to risk. This observation holds because we are assuming that, for all relevant μ ,

$$\hat{c}(\alpha, \mu) = \frac{\hat{c}(\alpha = 1, \mu)}{\alpha} < W.$$

When α is small enough

$$\frac{\hat{c}(\alpha = 1, \mu)}{\alpha}$$

can exceed W and in that case

$$\hat{c}(\alpha, \mu) = W$$

and the above argument fails to apply.

Since a more risk averse investor scales back his investment in the risky asset, the impact on utility of the optimal exposure to the excess return of the risky asset is

$$\alpha \hat{c}(\alpha, \mu)x = \hat{c}(1, \mu)x$$

for all levels of risk aversion α and for all possible distributions μ , whether they be the prior μ_0 or a posterior μ_s . This effectively means that the benefits of being informed are the same for all CARA investors regardless of how risk averse they are. More risk averse CARA investors with α higher do put a higher value on the cost of acquiring information.

Throughout the discussion, we have emphasized the fact that the investor's optimal investment in the risky asset is less than his wealth. If for some level of risk aversion this were not true it would be true for less risk averse investors. In that case, less risk averse investors would not, because of the wealth constraint invest more aggressively than, at least some, more risk averse investors and the Losq-Sobti argument would not apply. In the next section, we will focus on this issue explicitly.

108.5 The CRRA Case and the Simple Portfolio Problem

In this section, we return to the case in which the investor's utility function is in the CRRA class. In our earlier discussion of this case we focused on the complete market setting and the value of information when the investor chooses a portfolio of contingent claims in a complete market setting. In this section, we focus on the simple portfolio problem discussed by Losq and Sobti.

When an investor with relative risk aversion $\gamma > 0$ can invest in a safe and risky asset it is convenient to let c be the fraction of wealth invested in the risky asset. We will exposit the argument for the case in which $\gamma \neq 1$.

We will initially assume that the investor cannot sell short or borrow to invest and we have $p = 1$. So, initially, the constraint set is

$$C(W) = \{0 \leq c \leq 1\}.$$

At certain points in the discussion it will be important to consider the case in which the restrictions against short selling and borrowing will be ignored and c will be allowed to be negative or to exceed one. In fact, the cases in which the constraints are ignored play a central role in the discussion.

As in the CARA case, the return on the safe asset is one and the excess return on the risky asset is x . These assumptions imply that

$$\begin{aligned} \varphi(\gamma, W, \mu) &\equiv \varphi(u', W, \mu) \\ &= \max_{0 \leq c \leq 1} \frac{W^{1-\gamma}}{(1-\gamma)} \sum_{x \in X} (1+cx)^{1-\gamma} \mu(x) \\ &= W^{1-\gamma} \psi(\gamma, \mu) \end{aligned}$$

where

$$\psi(\gamma, \mu) = \max_{0 \leq c \leq 1} \sum_{x \in X} \frac{(1-cx)^{1-\gamma}}{(1-\gamma)} \mu(x)$$

is independent of W . When the borrowing and short selling prohibitions are ignored the same equations hold but c is not restricted to lie in the interval $[0,1]$.

The equation

$$E\varphi(\gamma, W(1-v), \mu_s^-) = \varphi(\gamma, W, \mu_0)$$

defining the value of information, $v = v(\gamma, E, \mu_0, \tilde{s})$ as a fraction of wealth now becomes

$$(1-v)^{1-\gamma} E\psi(\gamma, \mu_s^-) = \psi(\gamma, \mu_0).$$

So

$$v(\gamma, \mu_0, \tilde{s}) = 1 - \frac{[\psi(\gamma, \mu_0)]^{\frac{1}{1-\gamma}}}{[E\psi(\gamma, \mu_s)]^{\frac{1}{1-\gamma}}},$$

which is independent of wealth, W . Note that, in this case, the absolute risk aversion, γ/I , is decreasing in I while the value of information $v(\gamma, W, \mu_0, \tilde{s}) = v(\gamma, \mu_0, \tilde{s})W$ is increasing in W .

108.5.1 Some Examples

We now consider several simple examples in which there are only two possible excess returns. So $X = \{x_1, x_2\}$. Since we have two assets, this is a case of complete markets. We then introduce the possibility that the investor can observe a prediction, $s \in S = \{s_1, s_2\}$ of the return before making his portfolio choice. An observation of $\tilde{s} = s_i$ is interpreted as a prediction that $\tilde{x}$ will equal x_i in the sense that the probability of observing $\tilde{s} = s_i$ conditional on the return being x_i , $\Pr(s_i|x_i)$, exceeds one-half. For a range of relative risk aversion measures γ between 2.5 and 10.55 we compute $v(\gamma, \mu_0, \tilde{s})$, the fraction of wealth the investor would pay for the opportunity to make his portfolio choice after observing the prediction.

We let

$$X = \{x_1 = .24, x_2 = -.1\}$$

and

$$\mu_0(x_1) = .48.$$

So

$$(E_0\tilde{x}, \sigma_0\tilde{x}) = (.0632, .1699).$$

This mean and standard deviation are close to those reported for 1889–1978 in Table 1 of Mehra and Prescott (1985).

108.5.1.1 The First Example

In this example, the prediction is not particularly reliable in the sense that

$$\Pr(s_1|x_1) = \Pr(s_2|x_2) = .55.$$

The posterior when $\tilde{s} = s_1$ has been observed is

$$\langle \mu_{s_1}(x_1), \mu_{s_1}(x_2) \rangle = \langle .5301, .4699 \rangle.$$

So

$$(E[\tilde{x}|s_1], \sigma_{s_1}\tilde{x}) = (.0802, .1697).$$

Also, the posterior when $\tilde{s} = s_2$ has been observed is

$$\langle \mu_{s_2}(x_1), \mu_{s_2}(x_2) \rangle = \langle .4303, .5697 \rangle$$

So

$$(E[\tilde{x}|s_2], \sigma_{s_2}\tilde{x}) = (.0463, .1683).$$

Since the information provided by the prediction is not very reliable, the posterior moments are close to the prior moments. In the case of a prediction of a low return, the expected return remains positive and, as a consequence the investment in the risky asset, $\hat{c}$ also remains positive. The graph in Fig. 1 plots $v(\gamma, \mu_0, \tilde{s})$, for a range of relative risk aversion measures γ between 2.5 and 10.55.

The values of v are in the range of 0.04% to 0.15%. Most significantly, when γ is between 2.5 and 3.1, $v(\gamma, \mu_0, \tilde{s})$ is an increasing function of relative and, hence, absolute risk aversion. It turns out to be important to note that, for the γ values in this range, the investor holds only the risky asset when there is a prediction of a high return. So in that range, there is a corner solution.

To highlight the importance of the corner solution in this example, we can solve the portfolio problem allowing the investor to borrow and to choose c above one. In that case, he will indeed choose c to be above one when $\gamma < 3.1$ and there is a prediction of a high return. For this case, Fig. 2 plots $v(\gamma, \mu_0, \tilde{s})$, for the range of relative risk aversion measures γ between 2.5 and 10.55.

As illustrated in Fig. 2, we see that when c can be chosen to be above one, the value of information, $v(\gamma, \mu_0, \tilde{s})$, is everywhere a decreasing function of relative risk aversion and there is no range in which $v(\gamma, \mu_0, \tilde{s})$ is an increasing function of risk aversion.

108.5.1.2 The Second Example

In this example, the prediction is more reliable in the sense that

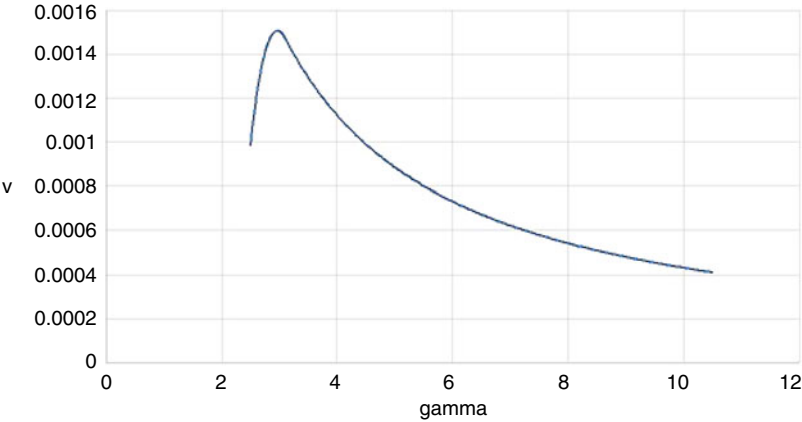


Fig. 1 First Example, No Borrowing or Short Sales

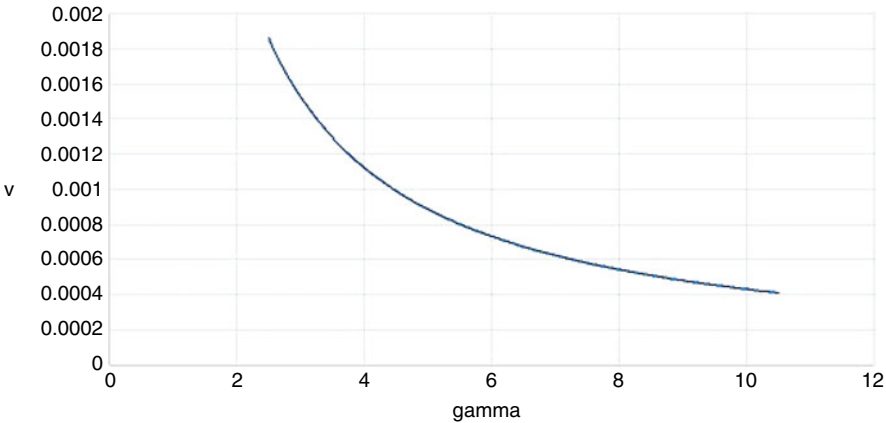


Fig. 2 First Example, Borrowing and Short Sales Permitted

$$\Pr(s_1|x_1) = \Pr(s_2|x_2) = .75.$$

The posterior when $\tilde{s} = s_1$ has been observed is

$$\langle \mu_{s_1}(x_1), \mu_{s_1}(x_2) \rangle = \langle .7347, .2653 \rangle.$$

So

$$(E[\tilde{x}|s_1], \sigma_{s_1}\tilde{x}) = (.1498, .1501).$$

Also

$$\langle \mu_{s_2}(x_1), \mu_{s_2}(x_2) \rangle = \langle .2353, .7647 \rangle.$$

So

$$(E[\tilde{x}|s_2], \sigma_{s_2}\tilde{x}) = (-.02, .1442).$$

In this case, the prediction of a low return implies a negative expected excess return and, as a consequence, any risk averse investor holds only the riskless asset. So there is a corner solution. Figure 3 plots $v(\gamma, \mu_0, \tilde{s})$, for a range of relative risk aversion measures γ between 2.5 and 10.55.

These v values are in the range of 1% to 2.6%. In this case, when γ is between 2.5 and 3.5, $v(\gamma, \mu_0, \tilde{s})$ is an increasing function of relative and, hence, absolute risk aversion. Again it should be noted that for the γ values in this range, the investor holds only the risky asset when there is a prediction of a high return. So again we have a corner solution.

Again, it turns out that the result that $v(\gamma, \mu_0, \tilde{s})$ is an increasing function of relative risk aversion for a range of γ 's depends crucially on the fact that, in that range, the investor's optimal portfolio is a corner solution. If we ignore the constraints and allow the investor to borrow and sell short, he will borrow when there is a prediction of a high return and sell short when the return prediction is low. In that case $v(\gamma, \mu_0, \tilde{s})$ is a decreasing function of relative risk aversion over the entire range of γ from 2.5 to 10.55. For this case, Fig. 4 plots $v(\gamma, \mu_0, \tilde{s})$, for the range of relative risk aversion measures γ between 2.5 and 10.55.

108.5.1.3 A Third Example

This example is motivated by the fact that, in the first two examples, the uninformed investors all invested positive amounts in the riskless asset and as risk aversion increased the fraction of wealth invested in the risky asset declined. In the final example, we raise the unconditional probability of $\tilde{x} = x_1$ to

$$\mu_0(x_1) = .6.$$

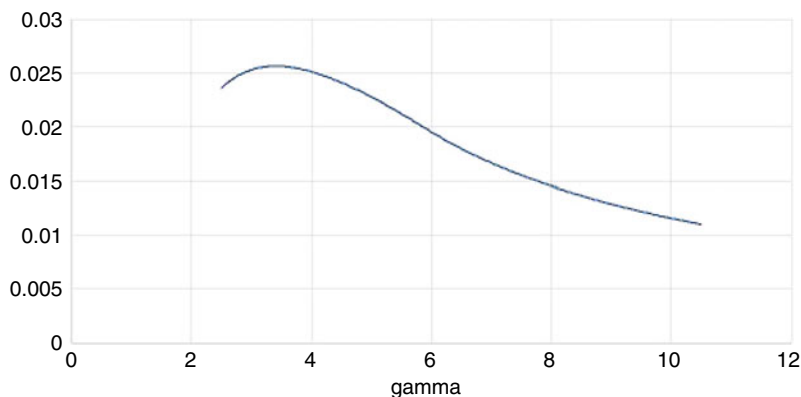


Fig. 3 Second Example, No Borrowing or Short Sales

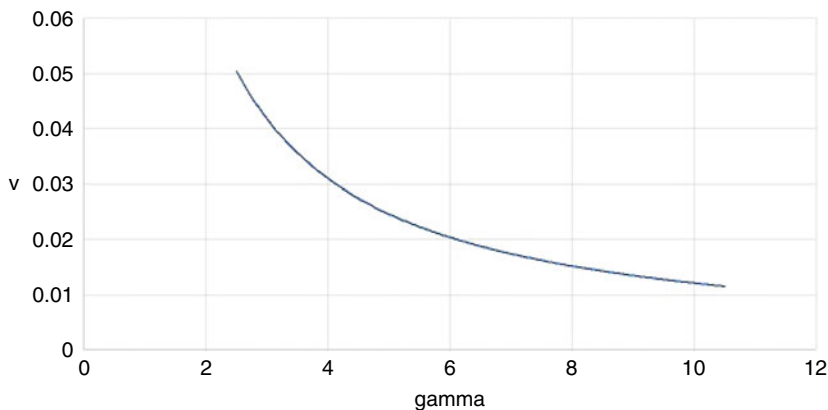


Fig. 4 Second Example, Borrowing and Short Sales Permitted

Then

$$(E_0\tilde{x}, \sigma_0\tilde{x}) = (.104, .1665).$$

In this case, the uninformed investors with γ between 2.5 and 3.95 all choose not to invest in the riskless asset. As risk aversion increases in this range these investors do not behave more cautiously and the fraction of wealth invested in the risky asset does not decline.

Again suppose that the prediction is more reliable in the sense that

$$\Pr(s_1|x_1) = \Pr(s_2|x_2) = .75$$

The posterior when $\tilde{s} = s_1$ has been observed is

$$\langle \mu_{s_1}(x_1), \mu_{s_1}(x_2) \rangle = \langle .82, .18 \rangle.$$

So

$$(E[\tilde{x}|s_1], \sigma_{s_1}\tilde{x}) = (.1782, .1311).$$

In response to this reliable good news, the investors with γ between 2.5 and 7.4 all choose not to invest in the riskless asset.

The posterior when $\tilde{s} = s_2$ has been observed is

$$\langle \mu_{s_2}(x_1), \mu_{s_2}(x_2) \rangle = \langle .33, .67 \rangle.$$

So

$$(E[\tilde{x}|s_2], \sigma_{s_2}\tilde{x}) = (.013, .1603).$$

In response to this reliable bad news, investors at all levels of γ between 2.5 and 10.55 choose to invest over 78% of their wealth in the riskless asset.

Figure 5 plots $v(\gamma, \mu_0, \tilde{s})$, for a range of relative risk aversion measures γ between 2.5 and 10.55. These v values are in the range of 0.9% to 2%. In this case, when γ is between 2.5 and 5.45, $v(\gamma, \mu_0, \tilde{s})$ is an increasing function of relative and, hence, absolute risk aversion. It should again be noted that for the γ values in this range, the investor holds only the risky asset when there is a prediction of a high return. The investors who choose not to invest in the riskless asset when they are uninformed, viz., those with γ between 2.5 and 3.95, are in this range.

Once again, it turns out that the result that $v(\gamma, \mu_0, \tilde{s})$ is an increasing function of relative risk aversion for a range of γ 's depends crucially on the fact that in that range the investor's optimal portfolio is a corner solution. If we ignore the constraints and allow the investor to borrow and sell short, he will when it is optimal. In that case $v(\gamma, \mu_0, \tilde{s})$ is a decreasing function of relative risk aversion. For this case, Fig. 6 plots $v(\gamma, \mu_0, \tilde{s})$, for the range of relative risk aversion measures γ between 2.5 and 10.55. Note that for the least risk averse investor $v = 5.3\%$ and for the most risk averse $v = 1.2\%$.

108.6 Concluding Remarks

The somewhat surprising result that a risk aversion increase reduces the value of information is rather general. Why does it happen? The answers are different when the uninformed investor chooses to face no risk and when he does choose to face risk. An uninformed investor who chooses not to bear risk and who can acquire information effectively has the opportunity to make a risky investment. More risk averse investors put a lower value on risky investment opportunities.

In the CARA case we considered, the uninformed investor does choose to face risk. In the CARA case, a more risk averse investor scales back his exposure to risk

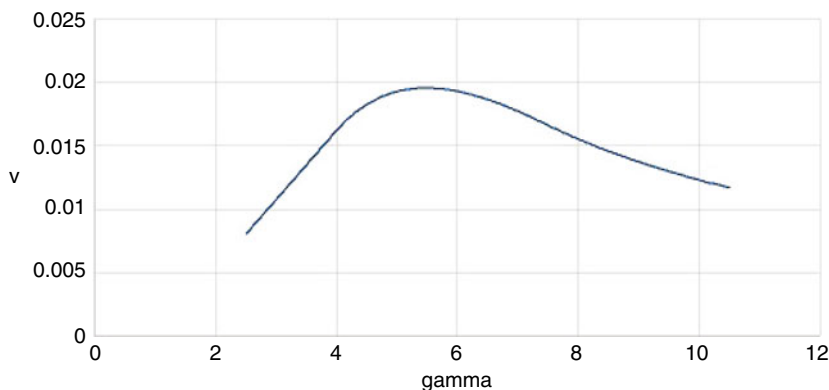


Fig. 5 Third Example, No Borrowing or Short Sales

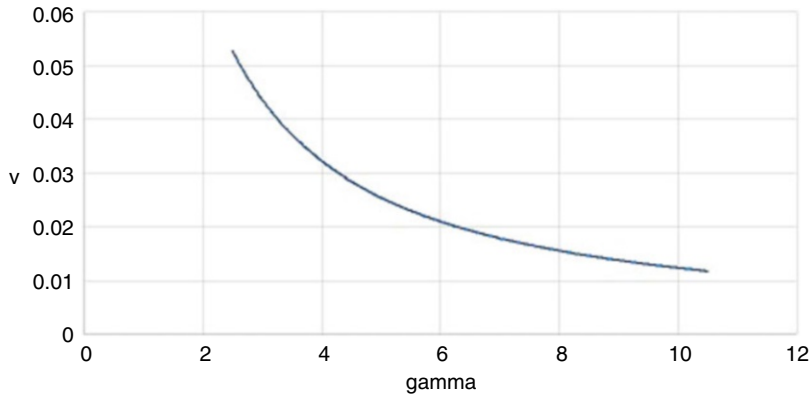


Fig. 6 Third Example, Borrowing and Short Sales Permitted

to such an extent that the impact of his increased risk aversion is completely eliminated. The more cautious behavior of the more risk averse investor means that he has no more need for information than the less risk averse investor. Caution effectively reduces the need for information. Alternatively, one could say that less risk averse investors use information more aggressively and that makes it more valuable.

In the simple CRRA examples for which we computed the value of information there were ranges of the relative risk aversion measure on which the value of information did rise with risk aversion. But in those ranges, the investor's optimal portfolio was constrained to be a corner solution. That effectively meant that increases in risk aversion did not imply more cautious behavior. The investor did not, as in the CARA case, scale back his investment in the risky asset as he became more risk averse. In all of the examples, when the borrowing and lending constraints were ignored, increases in risk aversion always implied more cautious optimal behavior on the part of the investor and the increases in risk aversion did lower the value of information. This suggests that, in a portfolio setting, when a more risk averse investor is able to invest more cautiously than a less risk averse investor, information will be less valuable. This observation is consistent with the gambling example considered by Gollier in his 2001 survey in Chap. 24 of Gollier (2001). In the Gollier example, the potential gambler either gambles or he does not. The size of the gamble is fixed. More risk averse gamblers cannot gamble less aggressively. In that example the least risk averse gamblers do gamble and the least risk averse among them put the lowest value on information.

References

- Arrow, K. 1965. "Aspects of the Theory of Risk Bearing," *The Theory of Risk Aversion*. Helsinki, Yrjö Johanssonin Saato. Reprinted in Arrow, K. 1971. *Essays in the Theory of Risk Bearing*, Markham, Chicago. 90–109.

- Blackwell, D. 1953. Equivalent comparisons of experiments. *Annals of Mathematical Statistics* 24: 265–272.
- Bonnenblust, Shapley, and Sherman. 1949. Reconnaissance in game theory, unpublished RAND Research Memorandum, RM-208.
- Cabrales, A., O. Gossner, and R. Serrano. 2013. Entropy and the value of information for investor. *American Economic Review* 103: 360–377.
- . 2017. A normalized value for information purchases. *Journal of Economic Theory* 170: 266–288.
- DeGroot, M. 1962. Uncertainty, information and sequential experiments. *Annals of Mathematical Statistics* 33: 404–419.
- Gollier, C. 2001. *The economics of risk and time*. MIT Press.
- Lindley, D.V. 1956. On a measure of the information provided by an experiment. *Annals of Mathematical Statistics* 27: 986–1005.
- Losq, E., and R. Sobti. 1985. Demand for information and risk aversion: Some results in a portfolio choice context, unpublished working paper, McGill University.
- Mehra, R., and E. Prescott. 1985. The equity premium puzzle. *Journal of Monetary Economics* 15: 145–161.
- Pratt, J. 1964. Risk aversion in the small and in the large. *Econometrica* 32: 122–136.



A Fuzzy Real Option Valuation Approach To Capital Budgeting Under Uncertainty Environment **109**

Shin-Yun Wang and Cheng-Few Lee

Abstract

The information needed for capital budgeting is generally not known with certainty. The sources of uncertainty may be the net cash inflows, the life of the project, or the discount rate. We propose a capital budgeting model under uncertainty environment in which the concept of probability is employed in describing fuzzy events and cash flow information can be specified as a special type of fuzzy numbers. The present worth of each fuzzy project cash flow can be subsequently estimated. At the same time, to select fuzzy projects and determine the optimal decision time under limited capital budget, we offer an example to analyze the results of the capital budgeting problem under uncertainty using a fuzzy real option valuation.

Keywords

Capital budgeting · Real option · Fuzzy numbers · Uncertainty · Fuzzy real option valuation

JEL Classification

D81 · G31

S.-Y. Wang (✉)

Department of Finance, National Dong Hwa University, Hualien, Taiwan
e-mail: gracew@gms.ndhu.edu.tw

C.-F. Lee

Department of Finance & Economics, Rutgers University, Piscataway, NJ, USA
e-mail: lee@business.rutgers.edu

109.1 Introduction

Large investments are capital projects of strategic importance which often have a long economic life cycle. They are characterized by many unknowns that are hard to foresee at their initial planning stage. Hence, the outcome of these investments may change during their long economic life, and the changes can be crucial since the larger the investments are, the more strategic importance they usually have. However, the information needed for capital budgeting is generally not known with certainty. Therefore, capital budgeting procedures under uncertainty remain to be further developed. The sources of uncertainty may be the net cash inflow, the life of the project, the cash flow volatility, and/or the discount rate.

In practice, precise information concerning future investment projects is rarely available. Hence, a decision maker does not have exact knowledge concerning future investment opportunities. Extensive studies have been undertaken to consider the capital budgeting problems under risk. Traditional approaches to capital budgeting are based on the premise that probability theory is necessary and sufficient to deal with the uncertainty that underlies the estimates of required parameters. However, it is argued that in many circumstances this premise is invalid since the principal sources of uncertainty are often nonrandom in nature and are related to the fuzziness rather than the frequency of data. In order to capture and quantify correctly the underlying uncertainty present in nonstatistical situations, the theory of possibility, an extension of the theory of fuzzy sets introduced by Zadeh (1965), can be used.

In the real world, the data sometimes cannot be recorded or collected precisely. For instance, if we ask a group of economists to predict the economic growth rate in the next year, their statements may be something like “approximately 3%,” “should be 3–5%,” “may be below 6%,” etc. All of these statements can be characterized as fuzzy sets. Suppose that we introduce a selection procedure which will determine the probability that each economist will be selected, then for each possible selection, their statements are fuzzy sets. Similarly, in a capital budgeting model under uncertainty, the cash flow information can be specified as a special type of fuzzy number (triangular fuzzy numbers). Zadeh (1965), Dubois and Prade (1988), and Carlsson and Fuller (2002) have investigated the usefulness of the fuzzy set theory in decision-making under uncertainty. Other studies involving decision-making use fuzzy modeling (Cabrer et al. 2009; Hu and Li 2009; Rao and Peng 2009), and fuzzy data mining (Hu and Tseng 2009; Peng et al. 2008). The fuzzy capital budgeting approaches allow cash flow estimates to be treated as fuzzy numbers, and in certain situations, fuzzy cash flows can better reflect the uncertainty in the project (Wang and Hwang 2010).

In this chapter, we integrate fuzzy versions of the neoclassical capital budgeting methods and real option valuation. In Sect. 109.2, we will discuss methods for capital budgeting and investment decision-making. Section 109.3 derives the fuzzy real option method. Section 109.4 gives a numerical example to compare and analyze the results of the capital budgeting problem using a fuzzy real option valuation. Conclusions are in Sect. 109.5.

109.2 Methods for Capital Budgeting and Investment Decision-Making

Capital budgeting methods based on the discounted cash flow (DCF) have been the primary instruments for investment decision-making. The most commonly used DCF-based method is the net present value (NPV). Under static circumstances, DCF-based methods provide reliable results. However, the real-world situations are seldom static. Especially in cases of large investments with long economic lives, the static DCF-based methods fail to present a highly reliable picture of the profitability and possibilities offered by the investment project. As DCF-based methods have been the best thing available, they have rooted to management practices during years of use.

Although there are many enhancements to the original formulae, the underlying unsatisfactory assumptions still exist. Hence, modified models are used which include the risk-adjusted discount-rate method and the certainty-equivalent method. The risk-adjusted discount-rate method simply extends the cash-flow valuation model under certainty to the uncertainty case. The advantage of this method is that the valuation model gives us a formula that explicitly considers the uncertainty associated with future cash flows. Although the model does not specify exactly what constitutes the risk of the cash flows, it can be used to develop and explore the relationships among the variables of asset-valuation models. On the other hand, the disadvantages of the risk-adjusted discount-rate model are clear. The value of risk-adjusted discount rate is only a subjective estimate, which could well differ from person to person. Therefore, an objective determination of the value of a risky investment project will be almost impossible by simply applying this method.

While the risk-adjusted discount-rate method provides a means for adjusting the basic riskless discount-rate, an alternative method, the certainty-equivalent method, adjusts the estimated value of the uncertain cash flows. The underlying rationale is that, given a risky cash flow, the decision maker will evaluate this risky cash flow by attaching an expected utility to that cash flow, which utility estimate being hypothesized to be equal to the utility derived from some certain amount. If a decision maker performs this process for each cash flow, a series of certainty equivalents for the risky flows can be obtained. These two models should yield the same value for a given stream of cash flows. However, Robichek and Myers (1966) show that the risk-adjusted discount-rate method tends to lump together the pure rate of interest, a risk premium, and time (through the compounding process), while the certainty-equivalent approach keeps risk and the pure rate of interest separate. This separation gives an advantage to the certainty-equivalent method.

Some economic factors, such as competition, inflation, consumer preferences, technological development, and labor market conditions are a few of them that make it impossible to foretell the future. Consequently, the economic life, revenues, and costs of investment projects are less than certain. With risk, a firm is no longer indifferent between two investment proposals having same internal rates of return or net present values. Three related stochastic methods have been useful in capital-budgeting decisions which are the probability distribution, decision tree, and simulation methods.

These methods have explicitly utilized the concepts of probability distributions and statistical distributions to carry out the analysis. If there is only a single accept-reject decision at the outset of the project, then a decision maker can use either statistical-distribution methods or simulation methods. If investment opportunities involve a sequence of decisions over time, a decision-tree method can be used to perform the analysis. Decision-tree is an analytical technique used in sequential decisions, where various decision points are studied in relation to subsequent chance events. This technique enables one to choose among alternatives in an objective and consistent manner. Simulation is another approach to confronting the problems of capital budgeting under uncertainty. Because uncertainty associated with capital budgeting is not restricted to one or two variables, every variable relevant in the capital-budgeting decision can be viewed as a random variable. Facing so many random variables, it may be impossible to obtain tractable results from an economic model. Simulation is a useful tool designed to deal with this problem and is the closest we can get to modeling in cases of uncertainty. However, if risk reduction is a consideration, then the real option is a better alternative. The real option approach also supports the determination of the timing for the investment and offers a comprehensive way of presenting a value of possibilities opened by the project, which can be further enhanced with frequently updated fuzzy cash flows.

The real option approach is a methodology that calculates the value of an investment with techniques originally developed for valuation of financial options. This gives the possibility to take into consideration the managerial flexibility to take action during the lifetime of an investment. The term real option was first introduced in an article about corporate borrowing by Myers and Turnbull (1977). Since then, there has been a growing literature to describe the different theoretical aspects of real options (Kulatilaka and Marcus 1988; Dixit and Pindyck 1994; Trigeorgis 1995), as well as the managerial and strategic implications and applications of real options (Bowman and Hurry 1993; Luehrman 1998; Amram and Kulatilaka 1999). A number of case-based articles are also available to give further insight into real-world application (Kulatilaka 1993; Nichols 1994; Micalizzi 1999).

109.3 A Fuzzy Real Option Valuation

109.3.1 Real Option Valuation

Real option means the possibility to either choose for or against making an investment decision without binding oneself up front. The real option rule is that one should invest today only if the net present value is high enough to compensate for giving up the value of the option to wait. While ongoing evaluation effort is structured around key decision points or triggered by changes in the business environment, it is a process in which the valuation is not intended to provide a definite answer, but rather to provide decision makers an ongoing dialogue about the project (Dahlberg and Porter 2000; Zeng and Zhang 2011).

In real options, the options involve real assets. Real options can be valued using the analogue option theories that have been developed for financial options, although they are not the same thing. Real options deal with strategic decisions of a company where degrees of freedom are limited to the capabilities of the company. In these strategic decisions, different stakeholders play a role, especially if the resources needed for an investment are significant and thereby the continuity of the company is at stake. Hence, it is quite different from traditional discounted cash flow investment approaches. Making decisions based upon the traditional methods is very hard when there is uncertainty about the exact outcome of the investment. In particular, since traditional methods ignore the value of flexibility and discount heavily for external uncertainty involved, many interesting innovative activities and projects are cancelled because of the uncertainties.

Thinking in real options is based on the Black-Scholes model in financial options. The option-pricing theory stipulates that one should invest today only if the net present value is high enough to compensate for giving up the value of the option to wait. Because the option to invest loses its value when the investment is irreversibly made, this loss is an opportunity cost of investing. In order to obtain the real option value (ROV) from the cash flow, the detail and inference process are explained below. Following the binomial model, we can compute the value of a real option as

$$ROV = F_0 B(n \geq a|T, p) - K(1+r)^{-T} B(n \geq a|T, p'), \quad (1)$$

where p is the cash flow up probability such that

$$p = \frac{(1+r-d)}{u-d}, p' = \frac{u}{1+r} p. \quad (2)$$

ROV denotes the current real option value; F_0 is the value of expected cash flows; $B(n \geq a|T, p)$ is the cumulative probability of having an in-the-money option (i.e., $n \geq a$), where the probabilities are the certainty-equivalent probabilities determined by the risk-free hedge portfolio; and $K(1+r)^{-T}$ is the discounted initial cash outflow. The option expires at time T , and r is the risk-free rate (see Copeland and Antikarov 2001 for details).

109.3.2 Probability of Fuzzy Events

The concept of probability is employed in describing fuzzy events and in using sample information to make statistical inferences (Lee et al. 2005a). An event is an experimental outcome that may or may not occur. The probability of a fuzzy event is used to measure the chance, the degree of compatibility, or degree of truth.

109.3.3 Prior Probability of Fuzzy Events

In the ROV (real option value) under uncertainty, the distributions of the primary variables are assessed subjectively. Therefore, a manager faces the problem of implicit fuzziness. It is difficult to measure the impreciseness with the concept of probability (Zadeh 1965; Tanaka and Asai 1984) because probability is used to measure randomness. Randomness is relevant to the occurrence or nonoccurrence of an event, while fuzziness is relevant to the degree of an event (Bellman and Zadeh 1970). According to the definition given by Zadeh (1965), a fuzzy set is used to describe the set of an event without clear boundaries. The membership function $\mu_{\tilde{A}} : X \rightarrow [0, 1]$ expresses the fuzzy set $\tilde{A}$ in set X where if element $x \in X$, then $\mu_{\tilde{A}}(x) \in [0, 1]$ expresses the grade of membership x (also, the degree of compatibility or degree of truth) of X in $\tilde{A}$, which maps X to the membership space. The greater the value of $\mu_{\tilde{A}}(x)$ is, the higher the grade of membership of x belongs to $\tilde{A}$. According to Zadeh (1968), when there is no extrasample information, the prior probability $P(\tilde{A})$ of fuzzy event $\tilde{A}$ can be defined as

$$P(\tilde{A}) = \sum_{i=1}^I \mu_{\tilde{A}}(x_i) \cdot P(x_i). \quad (3)$$

109.3.4 Posterior Probability of Fuzzy Events

Let $X = \{x_1, x_2, \dots, x_m\}$ be the sample information space. In S_k state, if the prior probability $P(x_j|S_k)$ of acquiring sample information x_j is known, then the posterior probability of acquiring sample information x_j is

$$P(S_k|x_j) = \frac{P(x_j|S_k) \cdot P(S_k)}{P(x_j)}. \quad (4)$$

One could evaluate the sample information acquired through this method. However, it still involves the subjective opinion of managers. On the other hand, the posterior probability of the sample information with fuzziness can be calculated as shown in the following. Let sample information space be $X = \{x_1, x_2, \dots, x_m\}$, let $\{x_r\}$, $r = 1, 2, \dots, m$ be an independent event, and let $\tilde{M} = \{\tilde{M}_1, \tilde{M}_2, \dots, \tilde{M}_J\}$ be the concept of fuzzy sample information. The posterior probability $P(S_k|\tilde{M}_j)$ is calculated in accordance with Bayes' rule after deriving $\tilde{M}_j$.

$$P(S_k|\tilde{M}_j) = \frac{P(\tilde{M}_j|S_k) \cdot P(S_k)}{P(\tilde{M}_j)},$$

where

$$P(\tilde{M}_j|S_k) = \sum_{r=1}^m P(x_r|S_k) \cdot \mu_{\tilde{M}_j}(x_r),$$

$$P(\tilde{M}_j) = \sum_{r=1}^m P(x_r) \cdot \mu_{\tilde{M}_j}(x_r),$$

therefore

$$P(S_k|\tilde{M}_j) = \frac{\sum_{r=1}^m P(x_r|S_k) \cdot \mu_{\tilde{M}_j}(x_r) \cdot P(S_k)}{\sum_{r=1}^m P(x_r) \cdot \mu_{\tilde{M}_j}(x_r)}. \quad (5)$$

When fuzzy sample information exists, the occurrence probability of state S_k can be described using the above formula. Therefore, the concept of “fuzzy” to describe the ROV under fuzzy environment should be brought into the picture.

When the range of the membership function is improved to $\{0, 1\}$, then $\tilde{A}$ will be transformed to a crisp set A . The $\mu_{\tilde{A}}$ will be transformed into C_A (characteristic function). This kind of transformation in the ROV means that the estimates of the primary variable cash flows, risk-rate, and the volatility of cash flow (F , R , σ) are without fuzziness, that is, $\mu_{\tilde{A}}(x_i) = 1$ or 0 . Under these circumstances, $\mu_{\tilde{A}}(x) = C_A(x) = 1$ if $x \in A$, or $\mu_{\tilde{A}}(x_i) = 0$, if $x \notin A$. Therefore, the probability for the occurrence of fuzzy event $\tilde{A}$ can be defined as

$$P(\tilde{A}) = \sum_{i=1}^I \mu_{\tilde{A}}(x_i) \cdot P(x_i) = \sum_{i=1}^I C_A(x_i) \cdot P(x_i) = P(A). \quad (6)$$

According to the traditional probabilistic ROV, the decision space faced by a manager is $B = \{S, D, P(S_i), ROV(d_j, S_i)\}$. Therefore, under state S_i and sample information x_r , the expected fuzzy real option value $\{FROV(d_j^*|x_r)\}$ of the optimal alternative is defined as

$$FROV(d_j^*|x_r) = \min_j \left\{ \sum_{i=1}^I ROV(d_j, S_i) \cdot P(S_i|x_r) \right\}$$

$$= \bigwedge_{j=1}^J \sum_{i=1}^I ROV(d_j, S_i) \cdot P(S_i|x_r). \quad (7)$$

When dealing with the actual ROV issues, a company not only faces a fuzzy sample information space but also stays in a fuzzy state space. For example, a manager forecasts future risk-free interest rate in accordance with the classification

of “booming economy,” “fair economy,” or “recession.” The definitions of “booming economy,” “fair economy,” and “recession” depend on the manager’s subjective opinion. Therefore, the state space encountered by the manager also involves implicit fuzziness. Besides, the actions that the manager plans to take will cause the cash flow structure to fluctuate. Moreover, due to the company’s environment, timing of the decision-making, and the inability to give it a trial, it is virtually impossible to sit and wait for a longtime for the collection of perfect information. Under these circumstances, the alternative adopted by the manager for ROV under uncertainty contains fuzziness. In summary, the ROV, which a manager actually deals with, is in a fuzzy state with fuzzy sample information and fuzzy action. As a result, the company can be defined with fuzzy decision space $\tilde{B} = \{\tilde{F}, \tilde{A}, P(\tilde{S}), ROV(\tilde{A}, \tilde{S})\}$, in which $\tilde{S} = \{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_K\}$ stands for fuzzy state set; $\tilde{S}_k, k = 1, 2, \dots, K$ stands for a fuzzy set in S ; and $S = \{S_1, S_2, \dots, S_I\}$ stands for the state set. $S_i, i = 1, 2, \dots, I$ stands for a state of the state set. $\tilde{A} = \{\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_N\}$ stands for fuzzy action set; $\tilde{A}_n, n = 1, 2, \dots, N$ stands for a fuzzy action set in D ; $D = \{d_1, d_2, \dots, d_J\}$ stands for action set; and $d_j, j = 1, 2, \dots, J$ stands for an action or alternative available for the company. $P(\tilde{S}_k)$ is the prior probability of $\tilde{S}_k$; $ROV(\tilde{A}, \tilde{S})$ is the evaluation function of $\tilde{A} \times \tilde{S}$. The prior probability of fuzzy state $P(\tilde{S}_k)$ is defined as

$$P(\tilde{S}_k) = \sum_{i=1}^I \mu_{\tilde{F}_k}(S_i) \cdot P(S_i). \quad (8)$$

Let $\tilde{M} = \{\tilde{M}_1, \tilde{M}_2, \dots, \tilde{M}_J\}$ be the fuzzy sample information space in X . $\tilde{M}_j, j = 1, 2, \dots, J$, is the fuzzy sample information; $X = \{x_1, x_2, \dots, x_m\}$ is the sample information space, and $\{x_r\}, r = 1, 2, \dots, m$, is an independent event.

The posterior probability of fuzzy state $\tilde{S}_k$ is defined in accordance with the posterior probability of fuzzy events after the fuzzy sample information $\tilde{M}_j$ is derived as

$$\begin{aligned} P(\tilde{S}_k | \tilde{M}_j) &= \sum_{i=1}^I \mu_{\tilde{S}_k}(S_i) \cdot P(S_i | \tilde{M}_j) \\ &= \sum_{i=1}^I \mu_{\tilde{S}_k}(S_i) \cdot \frac{\sum_{r=1}^m P(x_r | S_i) \cdot \mu_{\tilde{M}_j}(x_r) \cdot P(S_i)}{P(\tilde{M}_j)} \\ &= \frac{\sum_{i=1}^I \sum_{r=1}^m \mu_{\tilde{S}_k}(S_i) \cdot \mu_{\tilde{M}_j}(x_r) \cdot P(x_r | S_i) \cdot P(S_i)}{\sum_{r=1}^m \mu_{\tilde{M}_j}(x_r) \cdot P(x_r)}. \end{aligned} \quad (9)$$

109.3.5 A Fuzzy Real Option Valuation

In this section, we shall introduce a real option rule in a fuzzy setting (Carlsson and Fuller 2003; Dai et al. 2016; Su et al. 2019; Wang and Lee 2010; Ho and Liao 2011; Mao and Wu 2011), where the present value of expected cash flows is estimated by triangular fuzzy numbers. We will also determine the optimal exercise time by using the fuzzy real option valuation. To do so, we first briefly discuss the concept of triangular fuzzy number, followed by the combination of binomial model and triangular fuzzy numbers to form fuzzy decision tree.

As illustrated in Zadeh (1965), $\tilde{A}$ is a convex fuzzy set if and only if its α -level set $\tilde{A}_\alpha = \{x : \mu_{\tilde{A}}(x) \geq \alpha\}$ is a convex set for all α . Therefore, if $\tilde{a}$ is a fuzzy number, then the α -level set $\tilde{a}_\alpha$ is a compact (closed and bounded in $\mathbb{R}$) and convex set; that is, $\tilde{a}$ is a closed interval. The α -level set of $\tilde{a}$ is then denoted by $\tilde{a}_\alpha = [\tilde{a}_\alpha^L, \tilde{a}_\alpha^U]$. A fuzzy number $\tilde{a}$ is said to be nonnegative if $\mu_{\tilde{a}}(x) = 0$ for $x < 0$. It is easy to see that if $\tilde{a}$ is a nonnegative fuzzy number, then $\tilde{a}_\alpha^L$ and $\tilde{a}_\alpha^U$ are all nonnegative real numbers for all $\alpha \in [0, 1]$. The concept of triangular fuzzy number attempts to deal with real problems by possibility. The membership of a triangular fuzzy number is defined as follows (Fig. 1):

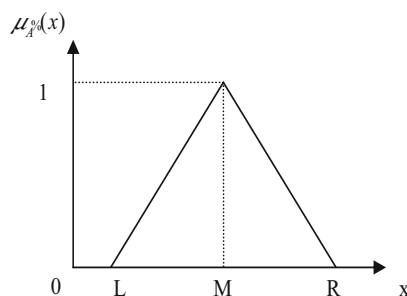
$$\mu_{\tilde{A}}(x) = \begin{cases} (x - L)/(M - L) \rightarrow L \leq x \leq M \\ (R - x)/(R - M) \rightarrow M \leq x \leq R \\ \rightarrow \rightarrow \rightarrow 0 \rightarrow \rightarrow \rightarrow \text{otherwise} \end{cases}$$

According to the characteristics of triangular fuzzy numbers and the extension principle put forward by Zadeh (1965), the operational laws of triangular fuzzy numbers, $\tilde{A} = (l_1, m_1, r_1)$ and $\tilde{B} = (l_2, m_2, r_2)$, are as follows:

1. Addition of two fuzzy numbers $\oplus$

$$(l_1, m_1, r_1) \oplus (l_2, m_2, r_2) = (l_1 + l_2, m_1 + m_2, r_1 + r_2)$$

Fig. 1 The membership function of the triangular fuzzy number



2. Subtraction of two fuzzy numbers $\ominus$

$$(l_1, m_1, r_1) \ominus (l_2, m_2, r_2) = (l_1 - r_2, m_1 - m_2, r_1 - l_2)$$

3. Multiplication of two fuzzy numbers $\otimes$

$$(l_1, m_1, r_1) \otimes (l_2, m_2, r_2) \cong (l_1 l_2, m_1 m_2, r_1 r_2)$$

4. Division of two fuzzy numbers $\oslash$

$$(l_1, m_1, r_1) \oslash (l_2, m_2, r_2) \cong (l_1/r_2, m_1/m_2, r_1/l_2)$$

Usually, the present value of expected cash flows cannot be characterized by a single number. But managers are able to estimate the present value of expected cash flows by using triangular possibility distribution of the form $\tilde{F} = [f_l, f_m, f_r]$, i.e., the most likely values of the present value of expected cash flows lie in the f_m (which is the core of the triangular fuzzy number $\tilde{F}$); f_r is the greatest value and f_l is the smallest value for the present value of expected cash flows.

In addition, the values are also influenced by the financial markets such as risk-free interest rate (discount rate) and the volatility of cash inflows. We use historical data to estimate them by using a triangular possibility distribution of the form $\tilde{R} = [r_l, r_m, r_r]$, $\tilde{\sigma} = [\sigma_l, \sigma_m, \sigma_r]$.

The most possible values of risk rate and the volatility of cash inflow lie in the interval r_m and σ_m (which is the core of the triangular fuzzy number $\tilde{R}$ and $\tilde{\sigma}$). r_r is the greatest value and r_l is the smallest value for discount rate, and σ_r is the greatest value and σ_l is the smallest value for the volatility of cash inflows.

The information needed for capital budgeting is generally not known with certainty. The sources of uncertainty may be the net cash inflows, the discount rate, the cash flow volatility, the life of the project, etc. Hence in addition to considering the fuzzy volatility, we also consider the fuzzy cash inflow $\tilde{F}$ and fuzzy interest rate $\tilde{R}$. Under these circumstances, we suggest the use of formulas described in the next section for computing fuzzy real option values.

109.4 N-Step Fuzzy Decision Tree

We now combine binomial model and triangular fuzzy number to form fuzzy decision tree (Lee et al. 2005b). Assume that decision tree will mature (project termination) in a given period and the cash flow is initially set to F ($t = 1$). Cash flows can move up or down in the next period ($t = 2$). There are three possible scenarios in the up movement, $F_{ur}(=ur \cdot F)$, $F_{um}(=um \cdot F)$, or $F_{ul}(=ul \cdot F)$, and three scenarios in the down movement, $F_{dr}(=dr \cdot F)$, $F_{dm}(=dm \cdot F)$, or $F_{dl}(=dl \cdot F)$ depending on volatility. Hence, in each time interval the cash flow moves from its initial value of F to one of the above six new values, F_{ur} , F_{um} , F_{ul} , F_{dr} , F_{dm} , or F_{dl} . These six values are defined as:

F_{ur} : cash flow of up movement under the greatest volatility, $F_{ur} = F_0 \cdot \left[e^{(1+\rho)\sigma\sqrt{\Delta t}} \right]$,

F_{um} : cash flow of up movement under medium volatility, $F_{um} = F_0 \cdot \left[e^{\sigma\sqrt{\Delta t}} \right]$,

F_{ul} : cash flow of up movement under the smallest volatility, $F_{ul} = F_0 \cdot \left[e^{(1-\rho)\sigma\sqrt{\Delta t}} \right]$,

F_{dr} : cash flow of down movement under the greatest volatility, $F_{dr} = F_0 \cdot \left[e^{-(1-\rho)\sigma\sqrt{\Delta t}} \right]$,

F_{dm} : cash flow of down movement under medium volatility, $F_{dm} = F_0 \cdot \left[e^{-\sigma\sqrt{\Delta t}} \right]$,

F_{dl} : cash flow of down movement under the smallest volatility, $F_{dl} = F_0 \cdot \left[e^{-(1+\rho)\sigma\sqrt{\Delta t}} \right]$.

The FROV in maturity are $R\tilde{O}V_{ur} = \max(F_{ur} - X, 0)$, $R\tilde{O}V_{um} = \max(F_{um} - X, 0)$, $R\tilde{O}V_{ul} = \max(F_{ul} - X, 0)$, $R\tilde{O}V_{dr} = \max(F_{dr} - X, 0)$, $R\tilde{O}V_{dm} = \max(F_{dm} - X, 0)$, and $R\tilde{O}V_{dl} = \max(F_{dl} - X, 0)$.

To illustrate the process of obtaining the greatest volatility, we use the replicating portfolio approach found in the financial options. Let

$$mF + B = ROV \quad (10)$$

where m is the number of units of the underlying asset under the greatest volatility and B the amount invested in bonds under the greatest volatility.

Assume we buy m units of underlying asset under the greatest volatility and invest amount B in risky assets (bonds), (r_l, r_m, r_h) ; after one period, the value of portfolio for up movement is: $mF_{ur} + B(1 + r_h) = R\tilde{O}V_{ur}$ and the one for down movement is: $mF_{dl} + B(1 + r_h) = R\tilde{O}V_{dl}$.

where ur is the range of the greatest volatility of "up" movement. The movement from F to $F_{ur}(=ur \cdot F)$ is therefore an "up" movement under the greatest volatility; dl is the range of the greatest volatility of "down" movement. The movement from F to $F_{dl}(=dl \cdot F)$ is therefore a "down" movement under the greatest volatility.

Suppose the value of the portfolio either under up or down movement is equal to the value of the corresponding FROV, we obtain

$$\begin{aligned} m &= \frac{C_{ur} - C_{dl}}{F_{ur} - F_{dl}}, B = \frac{C_{ur} - mF_{ur}}{(1 + r_h)} = \frac{C_{ur} - \left(\frac{C_{ur} - C_{dl}}{F_{ur} - F_{dl}} \right) F_{ur}}{(1 + r_h)} \\ &= \frac{\left(\frac{C_{ur}F_{ur} - C_{ur}F_{dl} - C_{ur}F_{ur} + C_{dl}F_{ur}}{F_{ur} - F_{dl}} \right)}{(1 + r_h)} = \frac{\left(\frac{C_{dl}F_{ur} - C_{ur}F_{dl}}{F_{ur} - F_{dl}} \right)}{(1 + r_h)}, \end{aligned}$$

where $P_u = \frac{(1+r_h)-dl}{ur-dl}$, $P_m = \frac{(1+r_m)-dm}{um-dm}$, $P_d = \frac{(1+r_l)-dr}{ul-dr}$.

The probabilities of three up movements are P_u , P_m , and P_d , while the probabilities of three down movement are $1-P_u$, $1-P_m$, and $1-P_d$, respectively. In the above

equations, $\tilde{R\hat{O}V}_u$, $\tilde{R\hat{O}V}_m$, and $\tilde{R\hat{O}V}_d$ are the right, medium, and the left values of the current fuzzy real option price, respectively. So we can get the FROV under the greatest volatility ($\tilde{R\hat{O}V}_u$) as: $\tilde{R\hat{O}V}_u = e^{-r_l \Delta t} [P_u \cdot \tilde{R\hat{O}V}_{ur} + (1 - P_u) \cdot \tilde{R\hat{O}V}_{dl}]$. Similarly, we also can get the FROV under medium and the smallest volatility as follows: $\tilde{R\hat{O}V}_m = e^{-r_m \Delta t} [P_m \cdot \tilde{R\hat{O}V}_{um} + (1 - P_m) \cdot \tilde{R\hat{O}V}_{dm}]$, $\tilde{R\hat{O}V}_d = e^{-r_h \Delta t} [P_d \cdot \tilde{R\hat{O}V}_{ul} + (1 - P_d) \cdot \tilde{R\hat{O}V}_{dr}]$. The process of fuzzy decision tree can be stated as follows:

- [Step 1]: Estimate ρ and σ value to compute the F_{ur} , F_{um} , F_{ul} , F_{dr} , F_{dm} , and F_{dl} . In other words, $\sigma(1 + \rho)$ is the volatility of the right value of up movement, $\sigma(1 - \rho)$ is the volatility of the left value of up movement, and $F_{ur}(=ur \cdot F)$, $F_{um}(=um \cdot F)$, and $F_{ul}(=ul \cdot F)$ are the degrees of up movement. In the down movement, $-\sigma(1 - \rho)$ is the volatility of the right value of down movement, $-\sigma(1 + \rho)$ is the volatility of the left value of down movement, and $F_{dr}(=dr \cdot F)$, $F_{dm}(=dm \cdot F)$, and $F_{dl}(=dl \cdot F)$ are the degrees of down movement.
- [Step 2]: Compute FROV from the fuzzy cash flow of $t = n$ to the FROV of $t = n$. We use the fuzzy cash flow at $t = n$ and the (nominal) value of fixed costs to compute the FROV. The process is the same with binomial tree model. $FROV_n = \max(F_n - X, 0)$, where F_n is the fuzzy cash flow at time t and X is the (nominal) value of fixed costs.
- [Step 3]: Compute the fuzzy real option price from $t = n$ to $t = n-1$. Now we integrate the fuzzy set with the binomial tree model and use different probabilistic for different volatilities and interest rates. Here, $\tilde{R\hat{O}V}_u$, $\tilde{R\hat{O}V}_m$, and $\tilde{R\hat{O}V}_d$ are the largest, middle, and the smallest values of the fuzzy option price, respectively. Each node of the fuzzy option price has three option prices, not just a single one, as in the binomial tree model. So we replace the fuzzy option prices $\tilde{R\hat{O}V}_{ur}$ and $\tilde{R\hat{O}V}_{dl}$ for the greatest volatility to get $\tilde{R\hat{O}V}_u = e^{-r_l \Delta t} [P_u \cdot \tilde{R\hat{O}V}_{ur} + (1 - P_u) \cdot \tilde{R\hat{O}V}_{dl}]$; $\tilde{R\hat{O}V}_{um}$ and $\tilde{R\hat{O}V}_{dm}$ for the medium volatility to get $\tilde{R\hat{O}V}_m = e^{-r_m \Delta t} [\tilde{R\hat{O}V}_{um} \cdot P_m + \tilde{R\hat{O}V}_{dm} \cdot (1 - P_m)]$; and $\tilde{R\hat{O}V}_{ul}$ and $\tilde{R\hat{O}V}_{dl}$ for the smallest volatility to get $\tilde{R\hat{O}V}_d = e^{-r_h \Delta t} [\tilde{R\hat{O}V}_{ul} \cdot P_d + \tilde{R\hat{O}V}_{dr} \cdot (1 - P_d)]$.
- [Step 4]: Compute the FROV from $t = n-1$ to $t = n-2$. We use the same principle as the calculation of the fuzzy option price from $t = n$ to $t = n-1$ with three fuzzy option prices at $t = n-1$. We still use the same way for the FROV from $t = n$ to $t = n-1$, and the value we get at every node will have multiple FROVs.
- [Step 5]: Compute the fuzzy option price from $t = n-2$ to $t = n-3, \dots, t = 1$. The same principle applies to the FROV from $t = n-1$ to $t = n-2$. For computing one FROV, we shall generalize the probabilistic decision rule for optimal investment strategy to a fuzzy setting: Here, T is the maximum deferral time, and t^* is the time at which investment (exercise the option) is made, $0 \leq t^* \leq T$, for which the option, $\tilde{R\hat{O}V}_{t^*}$, attains its maximum value $\tilde{R\hat{O}V}_{t^*} = \max_{t=0, 1, \dots, T} \tilde{R\hat{O}V}_t$.

However, to find a maximum element from the set $\{\tilde{R\hat{O}V}_0, \tilde{R\hat{O}V}_1, \dots, \tilde{R\hat{O}V}_T\}$ is not an easy task because it involves ranking of triangular fuzzy numbers. In our computerized implementation, we have employed the following function to

determine the expected fuzzy real option values $\{\tilde{R\ddot{O}V}_u, \tilde{R\ddot{O}V}_m, \tilde{R\ddot{O}V}_d\}$ of triangular form:

$$E(\tilde{R\ddot{O}V}_t) = \text{Max}\{[(\tilde{R\ddot{O}V}_u - \tilde{R\ddot{O}V}_d) + (\tilde{R\ddot{O}V}_m - \tilde{R\ddot{O}V}_d)]/3 + \tilde{R\ddot{O}V}_d, 0\},$$

where $\tilde{R\ddot{O}V}_u$, $\tilde{R\ddot{O}V}_m$, and $\tilde{R\ddot{O}V}_d$ are the largest, middle, and the smallest values of the fuzzy real option values, respectively. In the following subsections, we employ an example to describe the application of FROV under uncertainty.

109.4.1 Fuzzy State

Assume that the manager is projecting possible risk rates in the next term for different future outlooks of the economy: “booming,” “fair,” and “recession.” Suppose state set $S = \{S_1, S_2, S_3\}$, where S_i denotes the risk rate of ROV and the set represents a collection of fuzzy states. The probability of risk-rate state can be illustrated in Table 1. For example, when the economy is in the state of “booming,” the probability of a 3% risk rate is 0.6.

109.4.2 Fuzzy Sample Information

While F , K , T , and R can be derived directly from observation, the calculation of σ requires the use of cash flow over some past periods of time. In response to recent cash flow fluctuations, the manager has readjusted the magnitude of fluctuation for the cash flow (σ). Based on historical statistics, σ is revised downward from its current level of 60–40%. Suppose the sample message space $X = \{x_1, x_2, x_3\}$, where x_r , $r = 1, 2, 3$ denotes the cash flow and (x_r) is an independent event. The manager produces estimations based on two basic assumptions of “very high risk-rate growth” and “relatively flat risk-rate growth.” Hence the fuzzy sample message space can be expressed as $\tilde{M} = \{\tilde{M}_1, \tilde{M}_2\}$, where $\tilde{M}_1$ denotes very high risk-rate growth and $\tilde{M}_2$ a relatively flat risk-rate growth. It is also known that historically given S_i , company’s prior probability of the occurrence of x_r is $P(x_r | S_i)$, as shown in Table 2, and the prior probability of the membership function $u_{\tilde{M}_j}(x_r)$ and x_r for fuzzy sample message $\tilde{M}_j$, $j = 1, 2$ is $P(x_r)$, as shown in Table 3.

From the above data, we can get the prior probability and posterior probability shown in Table 4. Then we can compute and analyze the results of FROV, assuming that the initial outlay, K , is \$10,000 and the net cash flows are \$3000, \$5000, and \$7000 associated with each prior probability scenario. The probability and cash flow in fuzzy environment can be simplified as in Table 4.

In Table 4, we consider a project that requires an initial outlay (K) of \$10,000 (column 1). In each of the following time periods, the cash flow (columns 3, 5, and 7) to be received is not known with perfect certainty, but the probabilities associated with each cash flow in each period are assumed to be known (columns 2, 4, and 6),

Table 1 $\mu_{\tilde{F}_k}(S_i)$ and $P(S_i)$

	$S_i(R_i)$		
$\mu_{\tilde{F}_k}(S_i)/P(S_i)$	3%	4%	5%
$\mu_{\tilde{F}_1}(S_i)$	0.6	0.4	0
$\mu_{\tilde{F}_2}(S_i)$	0.2	0.7	0.1
$\mu_{\tilde{F}_3}(S_i)$	0.1	0.4	0.5
$P(S_i)$	0.3	0.5	0.2

Table 2 $P(x_r|S_i)$

		x_r		
S_i	R_i	\$3,000	\$5,000	\$7,000
S_1	3%	0.25	0.3	0.25
S_2	4%	0.5	0.4	0.5
S_3	5%	0.25	0.3	0.25

Table 3 $\mu_{\tilde{M}_j}(x_r)$ and $P(x_r)$

	x_r		
$\tilde{\mu}_{\tilde{M}_j}(x_r)/P(x_r)$	\$3,000	\$5,000	\$7,000
$\tilde{\mu}_{\tilde{M}_1}(x_r)$	0.3	0.2	0.25
$\tilde{\mu}_{\tilde{M}_2}(x_r)$	0.2	0.2	0.25
$P(x_r)$	0.5	0.4	0.5

so that we use triangular fuzzy number to deal with cash flow and probability. We note that columns (4) and (6) are fuzzy conditional-probability figures, where the later periods' expected cash flows depend highly on what occurred in earlier time periods. Given our fuzzy cash flow and fuzzy probability estimates for periods 1, 2, and 3, we find 27 possible fuzzy joint probabilities (column 8), each of which corresponds to a cash-flow series. Thus in the uppermost path, we find the fuzzy joint probability 0.015 being associated with a fuzzy cash flow of \$2000 in period 1, followed by fuzzy cash flows of \$2000 and \$1000 in periods 2 and 3, respectively. This particular path is the worst possible result; in nondiscounted dollar terms, there is a 50% loss on the investment (see Table 5 at $t = 1$). At the other end of the spectrum, we find the best possible outcome offers a return \$12,945 with a fuzzy joint probability of 0.0125 (see Table 5 at $t = 27$). The returns and fuzzy joint probabilities can be calculated similarly for the other 25 possible patterns.

While the primary hurdle in this process is in estimating the fuzzy cash flows and the fuzzy probabilities, we recognize that it is the fuzzy NPV figures we are ultimately interested in. Table 5 gives the fuzzy present value of the fuzzy cash flows and the fuzzy net present value of the project for each fuzzy cash-flow series, where a 4% discount rate was employed. When we multiply each joint probability by the expected net present value associated with that probability, we obtain the expected net present value for the project as a whole to be \$2,517.18.

Table 4 Illustration of conditional-probability distribution approach under uncertainty

(1) Initial outlay period 0	(2) Prior probability P (1)	(3) Net cash flow	(4) Posterior probability $P(2 1)$	(5) Fuzzy net cash flow	(6) Fuzzy conditional probability $P(3 2,1)$	(7) Fuzzy cash flow	(8) Fuzzy joint probability
					0.2	1000	0.015
			0.25	2000	0.5	2000	0.0375
					0.3	3000	0.0225
					0.2	2000	0.03
10,000	0.3	3000	0.5	3000	0.5	3000	0.075
					0.3	4000	0.045
					0.2	3000	0.015
			0.25	4000	0.5	4000	0.0375
					0.3	5000	0.0225
					0.3	2000	0.045
			0.3	3000	0.4	4000	0.06
					0.3	6000	0.045
					0.3	4000	0.06
	0.5	5000	0.4	5000	0.4	5000	0.08
					0.3	6000	0.06
					0.3	5000	0.045
			0.3	7000	0.4	7000	0.06

(continued)

Table 4 (continued)

(1) Initial outlay period 0	(2) Prior probability P (1)	(3) Net cash flow	(4) Posterior probability $P(2 1)$	(5) Fuzzy net <i>cash flow</i>	(6) Fuzzy conditional probability $P(3 2,1)$	(7) Fuzzy cash flow	(8) Fuzzy joint probability
					0.3	9000	0.045
					0.25	4000	0.0125
			0.25	5000	0.5	6000	0.025
					0.25	8000	0.0125
					0.25	5000	0.025
	0.2	7000	0.5	7000	0.5	7000	0.05
					0.25	9000	0.025
					0.25	7000	0.0125
			0.25	8000	0.5	9000	0.025
					0.25	11,000	0.0125

Based upon the cash flows and standard deviations from Table 5, we use triangular fuzzy number to obtain the $\tilde{F}$ and $\tilde{\sigma}(\tilde{F})$ in Table 6. We then use the discount rate $\tilde{R} = [0.03, 0.04, 0.05]$ to compute the FROV and transform them into the expected fuzzy real option value ($E(\tilde{ROV}_t)$).

Referring to Table 6, we are dealing with a case of uncertainty because the fuzzy probabilities associated with each fuzzy cash flow in each period are not known. According to Table 5, we can compute the standard deviation of fuzzy cash flows and use triangular fuzzy numbers to obtain the interval of fuzzy cash flows (column 2). Similarly, the computation of fuzzy variance of the present value of the expected cash flows uses the same process (column 3). Then based upon $\tilde{F}$ and $\tilde{\sigma}(\tilde{F})$, we can obtain column 4 (FROV). The fuzzy variance of the fuzzy cash flows and fuzzy real option values are triangular fuzzy numbers. Finally, we use the defuzzified fuzzy ranking method to determine the expected fuzzy real option values (column 5). From

Table 5 Fuzzy cash-flow and joint probability

(1) t	(2) Fuzzy cash-flow	(3) = (2)- K	(4) Fuzzy joint probability	(5) Fuzzy cash flow (σ)
1	4,661.2	-5,338.8	0.015	1021.913
2	5,550.2	-4,449.8	0.0375	1443.632
3	6,439.2	-3,560.8	0.0225	984.8824
4	6,474.76	-3,525.24	0.03	1131.085
5	7,363.76	-2,636.24	0.075	1544.94
6	8,252.76	-1,747.24	0.045	1008.12
7	8,288.32	-1,711.68	0.015	577.6831
8	9,177.32	-822.68	0.0375	741.2431
9	10,066.32	66.32	0.0225	440.8144
10	8,297.84	-1,602.16	0.045	998.557
11	10,175.84	175.84	0.06	693.0202
12	11,953.84	1,953.84	0.045	223.0023
13	12,024.96	2,024.96	0.06	240.0801
14	12,913.96	2,913.96	0.08	25.77347
15	13,802.96	3,802.96	0.06	195.4392
16	14,763.08	4,763.08	0.045	372.9275
17	16,541.08	6,541.08	0.06	866.1388
18	18,319.08	8,319.08	0.045	1127.269
19	13,948.04	3,948.04	0.0125	105.4258
20	15,726.04	5,726.04	0.025	430.2211
21	17,504.04	7,504.04	0.0125	502.9987
22	16,686.16	6,686.16	0.025	582.0294
23	18,464.16	8,464.16	0.05	1220.687
24	20242.16	10,242.16	0.025	1144.282
25	19,388.72	9,388.72	0.0125	713.7123
26	21,166.72	11,166.72	0.025	1290.468
27	22,944.72	12,944.72	0.0125	1111.285

Table 6 Expected fuzzy real option value

t	$\tilde{F}$	$\tilde{\sigma}(\tilde{F})$	$FROI$	$E(R\tilde{O}V_t)$
1	$\tilde{F}_1 = [2617, 4661, 6705]$	$\tilde{\sigma}_1 = [960, 1022, 1084]$	$\tilde{C}_1 = [-3765, -2532, -1097]$	0
2	$\tilde{F}_2 = [2663, 5550, 8437]$	$\tilde{\sigma}_2 = [1182, 1444, 1705]$	$\tilde{C}_2 = [-5856, -1899, 2721]$	0
3	$\tilde{F}_3 = [4469, 6439, 8409]$	$\tilde{\sigma}_3 = [920, 985, 1050]$	$\tilde{C}_3 = [-3387, -1480, 809]$	0
4	$\tilde{F}_4 = [4213, 6475, 8737]$	$\tilde{\sigma}_4 = [1005, 1131, 1257]$	$\tilde{C}_4 = [-4545, -1352, 2483]$	0
5	$\tilde{F}_5 = [4274, 7364, 10454]$	$\tilde{\sigma}_5 = [1119, 1545, 1970]$	$\tilde{C}_5 = [-6851, -744, 7922]$	109
6	$\tilde{F}_6 = [6237, 8253, 10269]$	$\tilde{\sigma}_6 = [906, 1008, 1110]$	$\tilde{C}_6 = [-3896, -473, 3729]$	0
7	$\tilde{F}_7 = [7133, 8288, 9444]$	$\tilde{\sigma}_7 = [531, 578, 624]$	$\tilde{C}_7 = [-3198, -408, 2833]$	0
8	$\tilde{F}_8 = [7695, 9177, 10660]$	$\tilde{\sigma}_8 = [731, 741, 752]$	$\tilde{C}_8 = [-1163, -25, 1368]$	60
9	$\tilde{F}_9 = [9185, 10066, 10948]$	$\tilde{\sigma}_9 = [343, 441, 539]$	$\tilde{C}_9 = [-5432, 295, 7906]$	923
10	$\tilde{F}_{10} = [6301, 8298, 10295]$	$\tilde{\sigma}_{10} = [901, 999, 1096]$	$\tilde{C}_{10} = [-3972, -266, 4316]$	26
11	$\tilde{F}_{11} = [8790, 10176, 11562]$	$\tilde{\sigma}_{11} = [656, 693, 730]$	$\tilde{C}_{11} = [-2140, 433, 3541]$	611
12	$\tilde{F}_{12} = [11508, 11954, 12400]$	$\tilde{\sigma}_{12} = [-8, 223, 454]$	$\tilde{C}_{12} = [842, 2152, 3717]$	2237
13	$\tilde{F}_{13} = [11545, 12025, 12505]$	$\tilde{\sigma}_{13} = [-19, 240, 499]$	$\tilde{C}_{13} = [770, 2196, 3899]$	2289
14	$\tilde{F}_{14} = [12862, 12914, 12966]$	$\tilde{\sigma}_{14} = [-394, 26, 446]$	$\tilde{C}_{14} = [6765, 2773, 9496]$	6345
15	$\tilde{F}_{15} = [13412, 13803, 14194]$	$\tilde{\sigma}_{15} = [-85, 195, 476]$	$\tilde{C}_{15} = [1619, 3313, 6015]$	3649

16	$\tilde{F}_{16} = [14017, 14763, 15509]$	$\tilde{\sigma}_{16} = [205, 373, 541]$	$\tilde{C}_{16} = [-5411, 337, 11259]$	2062
17	$\tilde{F}_{17} = [14809, 16541, 18723]$	$\tilde{\sigma}_{17} = [818, 866, 914]$	$\tilde{C}_{17} = [-1135, 2400, 7003]$	2756
18	$\tilde{F}_{18} = [16065, 18319, 20574]$	$\tilde{\sigma}_{18} = [975, 1127, 1280]$	$\tilde{C}_{18} = [-3213, 2656, 11864]$	3769
19	$\tilde{F}_{19} = [13737, 13948, 14159]$	$\tilde{\sigma}_{19} = [-43, 105, 254]$	$\tilde{C}_{19} = [1334, 3211, 5723]$	3423
20	$\tilde{F}_{20} = [14866, 15726, 16586]$	$\tilde{\sigma}_{20} = [323, 430, 537]$	$\tilde{C}_{20} = [-4289, 1435, 11019]$	2722
21	$\tilde{F}_{21} = [16498, 17504, 18510]$	$\tilde{\sigma}_{21} = [444, 503, 562]$	$\tilde{C}_{21} = [-2794, 2333, 9709]$	3083
22	$\tilde{F}_{22} = [15522, 16686, 17850]$	$\tilde{\sigma}_{22} = [523, 582, 641]$	$\tilde{C}_{22} = [-2547, 2122, 8664]$	2747
23	$\tilde{F}_{23} = [16023, 18464, 20906]$	$\tilde{\sigma}_{23} = [1018, 1221, 1423]$	$\tilde{C}_{23} = [-3453, 2263, 12298]$	3702
24	$\tilde{F}_{24} = [17954, 20242, 22531]$	$\tilde{\sigma}_{24} = [1025, 1144, 1263]$	$\tilde{C}_{24} = [-2345, 2840, 10969]$	3821
25	$\tilde{F}_{25} = [17961, 19389, 20816]$	$\tilde{\sigma}_{25} = [702, 714, 726]$	$\tilde{C}_{25} = [493, 2736, 5643]$	2957
26	$\tilde{F}_{26} = [18586, 21167, 23748]$	$\tilde{\sigma}_{26} = [1125, 1290, 1456]$	$\tilde{C}_{26} = [-2703, 2802, 12303]$	4134
27	$\tilde{F}_{27} = [20722, 22945, 25167]$	$\tilde{\sigma}_{27} = [1035, 1111, 1188]$	$\tilde{C}_{27} = [29493, 74345, 165301]$	4070

where t : time (in years), $\tilde{F}$: fuzzy cash flow, $\tilde{\sigma}(\tilde{F})$: variance of the present value expected fuzzy cash flows, $FROI$: fuzzy real option value, and $E(\tilde{ROV}_t)$: expected fuzzy real option value

Table 6, we know that the FROV will increase with increasing expected fuzzy cash flow estimates. If the fuzzy volatility of cash flow increases, then the FROV will also increase.

In sum, when the environment is uncertain, the manager will face a distribution of future cash flows rather than a single-point estimate of future cash flows in each period. To deal with these uncertain cash flows, we employ the fuzzy real option valuation method. Such valuation method recognizes the importance of uncertainty, which the NPV analysis assumes away. We use the fuzzy real option valuation to compare and analyze the project, then we obtain the expected fuzzy real option values for the project, $\tilde{R\ddot{O}V}_{t^*} = \max_{t=0,1,\dots,T} \tilde{R\ddot{O}V}_t$, where the maximum deferral time is $T = 27$ and the investment decision time (exercise the option) is time t^* ; $0 \leq t^* \leq T$. Generally, the longer the time to maturity, the greater will be the FROV. By using FROV model, we can get the optimal time to abandon the project. In Table 6, it is shown that the optimal time to abandon is $t = 14$ at which $E(\tilde{R\ddot{O}V}_t)$ attends its maximum value.

109.5 Conclusions

The fuzzy capital budgeting approaches allow the cash flows estimate to be fuzzy numbers. Fuzzy cash flows can better reflect the uncertainty in the project. The integration of qualitative information into the process gives better decision support. The fuzzy real option approach also supports the determination of the optimal time to abandon the project, which the traditional method cannot do. The FROV thus offers a comprehensive way of presenting the value of possibilities opened by the project, which is further enhanced with frequently updated fuzzy cash flows. Finally, we would like to stress that advanced decision methods such as real options and fuzzy capital budgeting open the chance to explore the value of flexibility inside and outside a project, and give further insight into the uncertainty of large investments. The uncertainty is genuine, i.e., we simply do not know the exact levels of future cash flows. Without introducing FROV, it would not be possible to formulate this genuine uncertainty. The proposed model that incorporates subjective judgments and statistical uncertainties may give managers a better understanding of the problem when making investment decisions.

The findings of this chapter and their implications for the managerial decision-making can be summarized as follows: (1) The FROV increases with increasing expected cash flows estimates; proactive management can influence this by developing market strategies to enhance the sales and trim the costs; (2) FROV increases with increasing fuzzy volatility; the corporate management can be proactive and find ways to expand to new markets and product innovations as end results of their strategic decisions; (3) the longer the time to maturity, the greater will be the FROV; proactive management can make sure of this development by maintaining protective barriers and maintaining a technological lead; (4) an increase in risk-free rate of returns will increase the FROV, and this can be further enhanced by closely

monitoring changes in interest rates; and (5) the FROV will decrease if value is lost during the postponement of the investment, but this can be countered by either creating business barriers for competitors or by better managing key resources.

Appendix Supplementary Explanation for Fuzzy Set

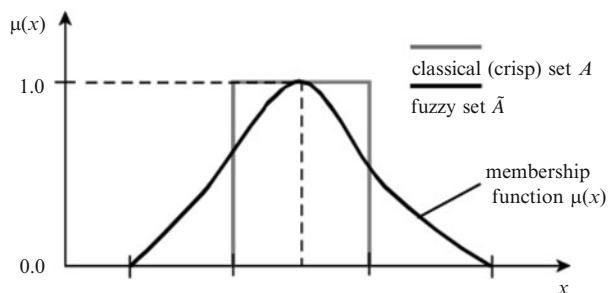
The Definition of Fuzzy Set

Fuzzy sets have been introduced by Zadeh (1965). What Zadeh proposed is very much a paradigm shift that first gained acceptance in the Far East, and its successful application has ensured its adoption around the world. Fuzzy sets are an extension of classical set theory and are used in fuzzy logic. In classical set theory, the membership of elements in relation to a set is assessed in binary terms according to a crisp condition – an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in relation to a set; this is described with the aid of a membership function valued in the real unit interval $[0, 1]$. Fuzzy sets are an extension of classical set theory since, for a certain universe, a membership function may act as an indicator function, mapping all elements to either 1 or 0, as in the classical notion. Specifically, a fuzzy set is any set that allows its members to have different grades of membership (membership function) in the interval $[0, 1]$. A fuzzy set on a classical set X is defined as follows:

$$\tilde{A} = \{(x, \mu_A(x)) | x \in X\}.$$

Membership Function

The membership function $\mu_A(x)$ quantifies the grade of membership of the elements x to the fundamental set X . An element mapping to the value 0 means that the member is not included in the given set; 1 describes a fully included member. Values strictly between 0 and 1 characterize the fuzzy members.



Fuzzy Logic

Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth values between “completely true” and “completely false.” As its name suggests, it is the logic underlying modes of reasoning which are approximate rather than exact. The importance of fuzzy logic derives from the fact that most modes of human reasoning and especially common sense reasoning are approximate in nature.

The Operations of Fuzzy Set

After knowing about the characteristic of fuzzy set, we will introduce the operations of fuzzy set. A fuzzy number is a convex, normalized fuzzy set whose membership function is at least segmental continuous and has the functional value $\mu_A(x) = 1$ at precisely one element. This can be likened to the funfair game “guess your weight,” where someone guesses the contestants weight, with closer guesses being more correct, and where the guesser “wins” if they guess near enough to the contestant’s weight, with the actual weight being completely correct (mapping to 1 by the membership function). A fuzzy interval is an uncertain set with a mean interval whose elements possess the membership function value $\mu_A(x) = 1$. As in fuzzy numbers, the membership function must be convex, normalized, and at least segmental continuous.

Defuzzify

When we are through the operations of fuzzy set to get the fuzzy interval, we will convert the fuzzy value into the crisp value. Below are some methods that convert a fuzzy set back into a single crisp (nonfuzzy) value. This is something that is normally done after a fuzzy decision has been made, and the fuzzy result must be used in the real world. For example, if the final fuzzy decision were to adjust the temperature setting on the thermostat a “little higher,” then it would be necessary to convert this “little higher” fuzzy value to the “best” crisp value to actually move the thermostat setting by some real amount.

Fuzzy Decision

After the process of defuzzifying, the next step is to make a fuzzy decision. Fuzzy decision is a model for decision-making in a fuzzy environment, and the object function and constraints are characterized as their membership functions, the intersection of fuzzy constraints, and fuzzy objection function.

References

- Amram, M., and N. Kulatilaka. 1999. *Real options managing strategic investments in an uncertain world*. Boston: Harvard Business School Press.
- Bellman, R.E., and L.A. Zadeh. 1970. Decision-making in a fuzzy environment. *Management Science* 17 (4): 141–164.
- Bowman, E.H., and D. Hurry. 1993. Strategy through the option lens: an integrated view of resource investments and incremental-choice process. *Academy of Management Review* 18: 760–842.
- Cabrer, F.J., S. Alonso, and E. Herrera-Viedma. 2009. A consensus model for group decision making problems with unbalanced fuzzy linguistic information. *International Journal of Information Technology and Decision Making* 8 (1): 109–131.
- Carlsson, C., and R. Fuller. 2002. *Fuzzy reasoning in decision making and optimization*. Heidelberg: Physica Verlag.
- . 2003. A fuzzy approach to real option valuation. *Fuzzy Sets and Systems* 139: 297–312.
- Copeland, T., and V. Antikarov. 2001. *Real options: a practitioner's guide*, 179–190. New York: W.W. Norton and Company.
- Dahlberg, K., and R.S. Porter. 2000. Get real. *Anderson Consulting Outlook Journal* 2: 33–40.
- Dai, H., T. Sun, and W. Guo. 2016. Brownfield redevelopment evaluation based on Fuzzy Real option. *Sustainability* 8 (2): 170–179.
- Dixit, A.K., and R.S. Pindyck. 1994. *Investment under uncertainty*. Princeton: Princeton University Press.
- Dubois, D., and H. Prade. 1988. *Possibility theory*. New York: Plenum Press.
- Ho, S.H., and S.H. Liao. 2011. A Fuzzy Real option approach for investment project valuation. *Expert Systems with Applications* 38 (12): 15296–15302.
- Hu, C., and S. Li. 2009. Two-phase interactive satisfying method of fuzzy multiple objective optimization with linguistic preference. *International Journal of Information Technology and Decision Making* 8 (3): 427–443.
- Hu, Y.-C., and F.-M. Tseng. 2009. Mining simplified fuzzy if-then rules for pattern classification. *International Journal of Information Technology and Decision Making* 8 (3): 473–489.
- Kulatilaka, N. 1993. The value of flexibility: the case of a dual-fuel industrial steam boiler. *Financial Management* 22 (3): 271.
- Kulatilaka, N., and A.J. Marcus. 1988. General formulation of corporate real options. *Research in Finance* 7: 183–282.
- Lee, C.F., G.H. Tzeng, and S.Y. Wang. 2005a. A new application of Fuzzy set theory to the Black-Scholes option pricing model. *Expert Systems with Applications* 29 (2): 330–342.
- . 2005b. A Fuzzy Set approach to generalize CRR model: An empirical analysis of S&P 500 index option. *Review of Quantitative Finance and Accounting* 25 (3): 255–275.
- Luehrman, T. 1998. Strategy as a portfolio of real options. *Harvard Business Review* 76: 89–99.
- Mao, Y., and W. Wu. 2011. Fuzzy Real Option evaluation of real estate project based on risk analysis. *Systems Engineering Procedia* 1: 228–235.
- Micalizzi, A. 1999. Timing to invest and value of managerial flexibility. Schering Plough case study (The 3rd Annual International Conference on Real Options, Wassenaar/Leiden).
- Myers, S.C., and S.M. Turnbull. 1977. Capital budgeting and the capital-asset pricing model: Good news and bad news. *Journal of Finance* 32: 321–333.
- Nichols, N.A. 1994. Scientific management at Merck. *Harvard Business Review* 72: 88–99.
- Peng, Y., G. Kou, Y. Shi, and Z. Chen. 2008. A descriptive framework for the field of data mining and knowledge discovery. *International Journal of Information Technology and Decision Making* 7 (4): 639–682.
- Rao, C., and J. Peng. 2009. Fuzzy group decision making model based on credibility theory and gray relative degree. *International Journal of Information Technology and Decision Making* 8 (3): 515–527.

- Robichek, A.A., and S.C. Myers. 1966. Conceptual problems in the use of risk-adjusted discount rates. *Journal of Finance* 12: 727–730.
- Su, C.S., T.Y. Chen, S.Y. Wang, and H.L. Wu. 2019. A study on alternative methods of real estate valuation: Fuzzy Real option model. *Journal of Housing Studies* 28 (1): 51–81.
- Tanaka, H., and K. Asai. 1984. Fuzzy linear programming problems with fuzzy numbers. *Fuzzy Sets and Systems* 13 (1): 1–10.
- Trigeorgis, L. 1995. *Real options in capital investment: models, strategies and applications*. Westport: Praeger.
- Wang, S.Y., and C.C. Hwang. 2010. An Application of Fuzzy Set theory to the weighted average cost of capital structure decision. *Technology and Investment* 1: 248–256.
- Wang, S.Y., and C.F. Lee. 2010. A Fuzzy Real option valuation approach to capital budgeting under uncertainty environment. *International Journal of Information Technology and Decision Making* 9 (2): 695–713.
- Zadeh, L.A. 1965. Fuzzy sets. *Information and Control* 8 (3): 338–353.
- . 1968. Probability measure of fuzzy events. *Journal of Mathematical Analysis and Applications* 23 (3): 421–427.
- Zeng, S., and S. Zhang. 2011. Real options literature review. *Scientific Research iBusiness* 3 (1): 43–48.

Appendix A: Derivation of Dividend Discount Model

A.1 Summation of Infinite Geometric Series

Summation of geometric series can be defined as:

$$S = A + AR + AR^2 + \cdots + AR^{n-1} \quad (\text{A.1})$$

Multiplying both sides of Equation A.1 by R , we obtain

$$RS = AR + AR^2 + \cdots + AR^{n-1} + AR^n \quad (\text{A.2})$$

Subtracting Equation A.1 by Equation A.2, we obtain

$$S - RS = A - AR^n$$

It can be shown

$$S = \frac{A(1 - R^n)}{1 - R} \quad (\text{A.3})$$

If R is smaller than 1, and n approaches to ∞ , then R^n approaches to 0 i.e.,

$$S_{\infty} = A + AR + AR^2 + \cdots + AR^{n-1} + \cdots + AR^{\infty}, \quad (\text{A.4})$$

then,

$$S_{\infty} = \frac{A}{1 - R} \quad (\text{A.5})$$

A.2 Dividend Discount Model

Dividend Discount Model can be defined as:

$$P_0 = \frac{D_1}{1+k} + \frac{D_2}{(1+k)^2} + \frac{D_3}{(1+k)^3} + \dots \quad (\text{A.6})$$

Where P_0 = present value of stock price per share D_t = dividend per share in period t ($t = 1, 2, \dots, n$)

If dividends grow at a constant rate, say g , then, $D_2 = D_1(1+g)$, $D_3 = D_2(1+g) = D_1(1+g)^2$, and so on.

Then, Equation A.6 can be rewritten as:

$$\begin{aligned} P_0 &= \frac{D_1}{1+k} + \frac{D_1(1+g)}{(1+k)^2} + \frac{D_1(1+g)^2}{(1+k)^3} + \dots \text{ or,} \\ P_0 &= \frac{D_1}{1+k} + \frac{D_1}{(1+k)} \times \frac{(1+g)}{(1+k)} + \frac{D_1}{(1+k)} \times \frac{(1+g)^2}{(1+k)^2} + \dots \end{aligned} \quad (\text{A.7})$$

Comparing Equation A.7 with Equation A.4, i.e., $P_0 = S_\infty$, $\frac{D_1}{1+k} = A$, and $\frac{1+g}{1+k} = R$ as in the Equation A.4.

Therefore, if $\frac{1+g}{1+k} < 1$ or if $k > g$, we can use Equation A.5 to find out P_0 i.e.,

$$\begin{aligned} P_0 &= \frac{D_1/(1+k)}{1 - [(1+g)/(1+k)]} \\ &= \frac{D_1/(1+k)}{[1+k - (1+g)]/(1+k)} \\ &= \frac{D_1/(1+k)}{(k-g)/(1+k)} \\ &= \frac{D_1}{k-g} = \frac{D_0(1+g)}{k-g} \end{aligned}$$

Appendix B: Derivation of DOL, DFL and DCL

B.1 DOL

Let P = price per unit

V = variable cost per unit

F = total fixed cost

Q = quantity of goods sold

The definition of DOL can be defined as: DOL (Degree of operating leverage)

$$\begin{aligned}
 &= \frac{\text{Percentage Change in Profits}}{\text{Percentage Change in Sales}} \\
 &= \frac{\Delta \text{EBIT} / \text{EBIT}}{\Delta \text{Sales} / \text{Sales}} \\
 &= \frac{\{[Q(P-V)-F] - [Q'(P-V)-F]\} / [Q(P-V)-F]}{(P \times Q - P \times Q') / (P \times Q)} \\
 &= \frac{[Q(P-V) - Q'(P-V)] / [Q(P-V)-F]}{P(Q-Q') / P \times Q} \\
 &= \frac{(Q-Q')(P-V) / [Q(P-V)-F]}{P(Q-Q') / P \times Q} \\
 &= \frac{\cancel{(Q-Q')}(P-V)}{Q(P-V)-F} \times \frac{P \times Q}{P \cancel{(Q-Q')}} \\
 &= \frac{Q(P-V)}{Q(P-V)-F} \\
 &= \frac{Q(P-V)-F+F}{Q(P-V)-F} = \frac{Q(P-V)-F}{Q(P-V)-F} + \frac{F}{Q(P-V)-F} \\
 &= 1 + \frac{F}{Q(P-V)-F} \\
 &= 1 + \frac{\text{Fixed Costs}}{\text{Profits}}
 \end{aligned}$$

B.2 DFL

$$\left. \begin{array}{l} \text{Let } i = \text{interest rate on} \\ \text{outstanding debt} \\ D = \text{outstanding debt} \end{array} \right\} \begin{array}{l} iD = \text{interest payment} \\ \text{on debt} \end{array}$$

N = the total number of shares outstanding

τ = corporate tax rate

$$EAIT = [Q(P - V) - F - iD](1 - \tau)$$

The definition of DFL can be defined as:

DFL (Degree of financial leverage)

$$\begin{aligned} &= \frac{\Delta \text{EPS/EPS}}{\Delta \text{EBIT/EBIT}} = \frac{(\Delta EAIT/N)/(\text{EAIT}/N)}{\Delta \text{EBIT/EBIT}} \\ &= \frac{\Delta \text{EAIT/EAIT}}{\Delta \text{EBIT/EBIT}} \\ &= \frac{[Q(P - V) - F - iD](1 - \tau) - [Q'(P - V) - F - iD]}{(1 - \tau)[Q(P - V) - F - iD](1 - \tau)} \\ &= \frac{[Q(P - V) - F] - [Q'(P - V) - F]}{[Q(P - V) - F]} \\ &= \frac{[Q(P - V)](1 - \tau) - [Q'(P - V)](1 - \tau)}{[Q(P - V) - F - iD](1 - \tau)} \\ &= \frac{[Q(P - V) - Q'(P - V)]}{[Q(P - V) - F]} \\ &= \frac{[\cancel{Q - Q'}(P - V)](\cancel{1 - \tau})}{[Q(P - V) - F - iD](\cancel{1 - \tau})} \times \frac{Q(P - V) - F}{\cancel{Q - Q'}(P - V)} \\ &= \frac{Q(P - V) - F}{Q(P - V) - F - iD} \left(= \frac{\text{EBIT}}{\text{EBIT} - iD} \right) \end{aligned}$$

B.3 DCL (Degree of Combined Leverage)

$$= \text{DOL} \times \text{DFL}$$

$$= \frac{Q(P - V)}{\cancel{Q(P - V)} - F} \times \frac{\cancel{Q(P - V)} - F}{Q(P - V) - F - iD} = \frac{Q(P - V)}{Q(P - V) - F - iD}$$

Appendix C: Derivation of Crossover Rate

Suppose there are two projects under consideration. Cash flows of project A, B and B – A are as follows:

Period	0	1	2	3
Project A	– 10,500	10,000	1,000	1,000
Project B	– 10,500	1,000	1,000	12,000
Cash flows of B – A	0	– 9,000	0	11,000

Based upon the information the table above we can calculate the NPV of Project A and Project B under different discount rates. The results are presented in Table A.1.

NPV(B) is higher with low discount rates and NPV(A) is higher with high discount rates. This is because the cash flows of project A occur early and those of project B occur later. If we assume a high discount rate, we would favor project A; if a low discount rate is expected, project B will be chosen. In order to make the right choice, we can calculate the crossover rate. If the discount rate is higher than the crossover rate, we should choose project A; if otherwise, we should go for project B.

The crossover rate, R_c , is the rate such that NPV(A) equals to NPV(B).

Suppose the crossover rate is R_c , then

$$\begin{aligned} \text{NPV}(A) = & -10,500 + 10,000/(1 + R_c) + 1,000/(1 + R_c)^2 \\ & + 1,000/(1 + R_c)^3 \end{aligned} \tag{A.8}$$

$$\begin{aligned} \text{NPV}(B) = & -10,500 + 1,000/(1 + R_c) + 1,000/(1 + R_c)^2 \\ & + 12,000/(1 + R_c)^3 \\ = & \text{NPV}(A) \end{aligned} \tag{A.9}$$

Therefore,

$$\begin{aligned} & -10,500 + \frac{10,000}{1 + R_c} + \frac{1,000}{(1 + R_c)^2} + \frac{1,000}{(1 + R_c)^3} \\ = & -10,500 + \frac{1,000}{1 + R_c} + \frac{1,000}{(1 + R_c)^2} + \frac{12,000}{(1 + R_c)^3} \end{aligned}$$

Table A.1 NPV of Project A and B under different discount rates

Discount rate (%)	NPV (Project A)	NPV (Project B)
0	1500.00	3500.00
5	794.68	1725.46
10	168.67	251.31
15	−390.69	−984.10
20	−893.52	−2027.78

Rearranging the above equation (moving all terms on the LHS to the RHS), we obtain Equation A.10

$$\begin{aligned} 0 = & [-10,500 - (-10,500)] + \left[\frac{1,000}{1 + R_c} - \frac{10,000}{1 + R_c} \right] \\ & + \left[\frac{1,000}{(1 + R_c)^2} - \frac{1,000}{(1 + R_c)^2} \right] + \left[\frac{12,000}{(1 + R_c)^3} - \frac{1,000}{(1 + R_c)^3} \right] \end{aligned} \tag{A.10}$$

Solving Equation A.10 by trial and error method for R_c , R_c equals 10.55%.
Using the procedure of calculating internal rate of return (IRR) as discussed in Equations A.8, A.9, and A.10, we calculate the IRR for both Project A and Project B. The IRR for Project A and B are 11.45% and 10.95% respectively. From this information, we have concluded that Project A will perform better than Project B without consideration for change of discount rate. Therefore, the IRR decision rule cannot be used for capital budgeting decisions when there exists an increasing or decreasing net cash inflow. This is so called “The Timing Problem” for using the IRR method for capital budgeting decisions.

Appendix D: Capital Budgeting Decisions with Different Lives

D.1 Mutually Exclusive Investment Projects with Different Lives

The traditional NPV technique may not be the appropriate criterion to select a project from mutually exclusive investment projects, if these projects have different lives. The underlying reason is that, compared with a long-life project, a short-life project can be replicated more quickly in the long run. In order to compare projects with different lives, we compute the NPV of an infinite replication of the investment project. For example, let Projects A and B be two mutually exclusive investment projects with the following cash flows.

Year	Project A	Project B
0	100	100
1	70	50
2	70	50
3		50

By assuming a discount rate of 12%, the traditional NPV of Project A is 18.30 and the NPV of Project B is 20.09. This shows that Project B is a better choice than Project A. However, the NPV with infinite replications for Project A and B should be adjusted into a comparable basis.

In order to compare Projects A and B, we compute the NPV of an infinite stream of constant scale replications. Let $NPV(N, \infty)$ be the NPV of an N -year project with NPV (N), replicated forever. This is exactly the same as an annuity paid at the beginning of the first period and at the end of every N years from that time on. The NPV of the annuity is:

$$NPV(N, \infty) = NPV(N) + \frac{NPV(N)}{(1 + K)^N} + \frac{NPV(N)}{(1 + K)^{2N}} + \cdots$$

In order to obtain a closed-form formula, let $(1/[(1 + K)^N]) = H$. Then we have:

$$NPV(N, t) = NPV(N)(1 + H + H^2 + \cdots + H^t) \tag{A.11}$$

Multiplying both sides by H , this becomes

$$H[NPV(N, t)] = NPV(N)(H + H^2 + \dots + H^t + H^{t+1}) \quad (\text{A.12})$$

Subtracting Equation A.12 from Equation A.11 gives:

$$\begin{aligned} NPV(N, t) - (H)NPV(N, t) &= NPV(N)(1 - H^{t+1}) \\ NPV(N, t) &= \frac{NPV(N)(1 - H^{t+1})}{1 - H} \end{aligned}$$

Taking the limit as the number of replications, t , approaches infinity gives:

$$\begin{aligned} \lim_{t \rightarrow \infty} NPV(N, t) &= NPV(N, \infty) \\ &= NPV \left[\frac{1}{1 - [1/(1 + K)^N]} \right] \\ &= NPV(N) \left[\frac{(1 + K)^N}{(1 + K)^N - 1} \right] \end{aligned} \quad (\text{A.13})$$

Equation A.13 is the NPV of an N -year project replicated at constant scale an infinite number of times. We can use it to compare projects with different lives because when their cash-flow streams are replicated forever, it is as if they had the same (infinite) life.

Based upon Equation A.13, we can calculate the NPV of Projects A and B as follows:

For Project A

$$\begin{aligned} &NPV(2, \infty) \\ &= NPV(2) \left[\frac{(1 + 0.12)^2}{(1 + 0.12)^2 - 1} \right] \\ &= (18.30) \left[\frac{1.2544}{0.2544} \right] \\ &= 90.23 \end{aligned}$$

For Project B

$$\begin{aligned} &NPV(3, \infty) \\ &= NPV(3) \left[\frac{(1 + 0.12)^3}{(1 + 0.12)^3 - 1} \right] \\ &= 20.09 \left[\frac{1.4049}{0.4049} \right] \\ &= 69.71 \end{aligned}$$

Consequently, we would choose to accept Project A over Project B, because, when the cash flows are adjusted for different lives, A provides the greater cash flow.

Alternatively, Equation A.13 can be rewritten as an equivalent annual NPV version as:

$$K \times NPV(N, \infty) = \frac{NPV(N)}{\text{Annuity factor}} \quad (\text{A.14})$$

where the annuity factor is

$$\frac{1 - 1/(1 + K)^N}{K}$$

The decision rule from Equation A.14 is equivalent to the decision rule of Equation A.13.

The different project lives can affect the beta coefficient estimate, as shown by Meyers and Turnbull (1977). For empirical guidance for evaluating capital-investment alternatives with unequal lives, the readers are advised to refer Emery (1982).

D.2 Equivalent Annual Cost

Equation A.14 can be written as:

$$NPV(N) = K \times NPV(N, \infty) \times \text{Annuity Factor} \quad (\text{A.15})$$

Corporate Finance by Ross, Westerfield, and Jaffe (2005, 7th edn, p. 193) has discussed about Equivalent Annual Cost. The Equivalent Annual Cost (C) can be calculated as follows:

$$NPV(N) = C \times \text{Annuity Factor} \quad (\text{A.16})$$

From Equations A.15 and A.16, we obtain

$$C = K \times NPV(N, \infty) \quad (\text{A.17})$$

Assume company A buys a machine that costs \$1,000 and the maintenance expense of \$250 is to be paid at the end of each of the 4 years. To evaluate this investment, we can calculate the present value of the machine. Assuming the discount rate as 10%, we have

$$\begin{aligned} NPV(A) &= 1000 + \frac{250}{1.1} + \frac{250}{(1.1)^2} + \frac{250}{(1.1)^3} + \frac{250}{(1.1)^4} \\ &= 1792.47 \end{aligned} \quad (\text{A.18})$$

Equation A.18 shows that payments of (1,000, 250, 250, 250, 250) are equivalent to a payment of 1792.47 at time 0. Using Equation A.16, we can equate the payment at time 0 of 1792.47 with a 4 year annuity.

$$\begin{aligned} 1792.47 &= C \times A_{0.1}^4 = C \times 3.1699 \\ C &= 565.47 \end{aligned}$$

In this example, following Equation A.13, we can find

$$\begin{aligned} NPV(N, \infty) &= 1749.47 \times (1 + 0.1)^4 / [(1 + 0.1)^4 - 1] \\ &= 5654.71 \end{aligned}$$

Then following the Equation A.17, we obtain

$$C = K \times NPV(N, \infty) = 0.1 \times 5654.71 = 565.47$$

Therefore, the equivalent annual cost C is identical to the equivalent annual NPV as defined in Equation A.14.

Appendix E: Derivation of Minimum-Variance Portfolio

If there is a two security portfolio, its variance can be defined as:

$$\sigma_p^2 = w_D^2 \sigma_D^2 + w_E^2 \sigma_E^2 + 2w_D w_E \text{Cov}(r_D, r_E) \quad (\text{A.19})$$

where r_D and r_E are the rate of return for security D and security E respectively; w_D and w_E are weight associated with security D and E respectively; σ_D^2 and σ_E^2 are variance of security D and E respectively; and $\text{Cov}(r_D, r_E)$ is the covariance between r_D and r_E .

The problem is choosing optimal w_D to minimize the portfolio variance, σ_p^2

$$\underset{w_D}{\text{Min}} \sigma_p^2 \quad (\text{A.20})$$

We can solve the minimization problem by differentiating the σ_p^2 with respect to w_D and setting the derivative equal to 0 i.e., we want to solve

$$\frac{\partial \sigma_p^2}{\partial w_D} = 0 \quad (\text{A.21})$$

Since, $w_D + w_E = 1$ or, $w_E = 1 - w_D$ therefore, the variance, σ_p^2 , can be rewritten as

$$\begin{aligned} \sigma_p^2 &= w_D^2 \sigma_D^2 + w_E^2 \sigma_E^2 + 2w_D w_E \text{Cov}(r_D, r_E) \\ &= w_D^2 \sigma_D^2 + (1 - w_D)^2 \sigma_E^2 + 2w_D(1 - w_D) \text{Cov}(r_D, r_E) \\ &= w_D^2 \sigma_D^2 + \sigma_E^2 - 2w_D \sigma_E^2 + w_D^2 \sigma_E^2 + 2w_D \text{Cov}(r_D, r_E) - 2w_D^2 \text{Cov}(r_D, r_E) \end{aligned}$$

Now, the first order conditions of Equation A.21 can be written as

$$2w_D \sigma_D^2 - 2\sigma_E^2 + 2w_D \sigma_E^2 + 2 \text{Cov}(r_D, r_E) - 4w_D \text{Cov}(r_D, r_E) = 0$$

Rearranging the above equation,

$$\begin{aligned}w_D\sigma_D^2 + w_D\sigma_E^2 - 2w_D\text{Cov}(r_D, r_E) &= \sigma_E^2 - \text{Cov}(r_D, r_E) \\[\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)]w_D &= \sigma_E^2 - \text{Cov}(r_D, r_E)\end{aligned}$$

Finally, we have

$$w_D = \frac{\sigma_E^2 - \text{Cov}(r_D, r_E)}{\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)}$$

Appendix F: Derivation of an Optimal Weight Portfolio Using the Sharpe Performance Measure

Solution for the weights of the optimal risky portfolio can be found by solving the following maximization problem:

$$\underset{w_D}{\text{Max}} S_p = \frac{E(r_p) - r_f}{\sigma_p}$$

where $E(r_p)$ = expected rates of return for portfolio P

r_f = risk free rates of return

S_p = sharpe performance measure, and

σ_p as defined in Equation A.19 of Appendix 5

We can solve the maximization problem by differentiating the S_p with respect to w_D , and setting the derivative equal to 0 i.e., we want to solve

$$\frac{\partial S_p}{\partial w_D} = 0 \quad (\text{A.22})$$

In the case of two securities, we know that

$$E(r_p) = w_D E(r_D) + w_E E(r_E) \quad (\text{A.23})$$

$$\sigma_p = [w_D^2 \sigma_D^2 + w_E^2 \sigma_E^2 + 2w_D w_E \text{Cov}(r_D, r_E)]^{1/2} \quad (\text{A.24})$$

$$w_D + w_E = 1 \quad (\text{A.25})$$

From above Equations A.23, A.24, and A.25, we can rewrite $E(r_p) - r_f$ and σ_p as:

$$\begin{aligned} E(r_p) - r_f &= w_D E(r_D) + w_E E(r_E) - r_f \\ &= w_D E(r_D) + (1 - w_D) E(r_E) - r_f \\ &\equiv f(w_D) \end{aligned} \quad (\text{A.26})$$

$$\begin{aligned}
\sigma_p &= [w_D^2 \sigma_D^2 + w_E^2 \sigma_E^2 + 2w_D w_E \text{Cov}(r_D, r_E)]^{1/2} \\
&= [w_D^2 \sigma_D^2 + (1 - w_D)^2 \sigma_E^2 + 2w_D(1 - w_D) \text{Cov}(r_D, r_E)]^{1/2} \\
&\equiv g(w_D)
\end{aligned} \tag{A.27}$$

Equation A.22 becomes

$$\begin{aligned}
\frac{\partial S_p}{\partial w_D} &= \frac{\partial [f(w_D)/g(w_D)]}{\partial w_D} \\
&= \frac{f'(w_D)g(w_D) - f(w_D)g'(w_D)}{[g(w_D)]^2} = 0
\end{aligned} \tag{A.28}$$

$$\text{where } f'(w_D) = \frac{\partial f(w_D)}{\partial w_D} = E(r_D) - E(r_E) \tag{A.29}$$

$$\begin{aligned}
g'(w_D) &= \frac{\partial g(w_D)}{\partial w_D} \\
&= \frac{1}{2} \times [w_D^2 \sigma_D^2 + (1 - w_D)^2 \sigma_E^2 + 2w_D(1 - w_D) \text{Cov}(r_D, r_E)]^{1/2-1} \\
&\quad \times [2w_D \sigma_D^2 + 2w_D \sigma_E^2 - 2\sigma_E^2 + 2\text{Cov}(r_D, r_E) - 4w_D \text{Cov}(r_D, r_E)] \\
&= [w_D \sigma_D^2 + w_D \sigma_E^2 - \sigma_E^2 + \text{Cov}(r_D, r_E) - 2w_D \text{Cov}(r_D, r_E)] \\
&\quad \times [w_D^2 \sigma_D^2 + (1 - w_D)^2 \sigma_E^2 + 2w_D(1 - w_D) \text{Cov}(r_D, r_E)]^{-1/2}
\end{aligned} \tag{A.30}$$

From Equation A.28,

$$\begin{aligned}
f'(w_D)g(w_D) - f(w_D)g'(w_D) &= 0, \text{ or} \\
f'(w_D)g(w_D) &= f(w_D)g'(w_D)
\end{aligned} \tag{A.31}$$

Now, plugging $f(w_D)$, $g(w_D)$, $f'(w_D)$, and $g'(w_D)$ [Equations A.26, A.27, A.29, and A.30] into Equation A.31, we have

$$\begin{aligned}
&[E(r_D) - E(r_E)] \\
&\quad \times [w_D^2 \sigma_D^2 + (1 - w_D)^2 \sigma_E^2 + 2w_D(1 - w_D) \text{Cov}(r_D, r_E)]^{1/2} \\
&= [w_D E(r_D) + (1 - w_D) E(r_E) - r_f] \\
&\quad \times [w_D \sigma_D^2 + w_D \sigma_E^2 - \sigma_E^2 + \text{Cov}(r_D, r_E) - 2w_D \text{Cov}(r_D, r_E)] \\
&\quad \times [w_D^2 \sigma_D^2 + (1 - w_D)^2 \sigma_E^2 + 2w_D(1 - w_D) \text{Cov}(r_D, r_E)]^{-1/2}
\end{aligned} \tag{A31}$$

Multiplying by $[w_D^2 \sigma_D^2 + (1 - w_D)^2 \sigma_E^2 + 2w_D(1 - w_D) \text{Cov}(r_D, r_E)]^{1/2}$ on both sides of Equation A.32, we have

$$\begin{aligned}
& [E(r_D) - E(r_E)] \\
& \times \left[w_D^2 \sigma_D^2 + (1 - w_D)^2 \sigma_E^2 + 2w_D(1 - w_D)\text{Cov}(r_D, r_E) \right] \\
& = [w_DE(r_D) + (1 - w_D)E(r_E) - r_f] \\
& \times [w_D \sigma_D^2 + w_D \sigma_E^2 - \sigma_E^2 + \text{Cov}(r_D, r_E) - 2w_D \text{Cov}(r_D, r_E)]
\end{aligned} \tag{A.33}$$

Rearrange all terms on both hand sides of Equation A.33, i.e.,

Left hand side of Equation A.33

$$\begin{aligned}
& [E(r_D) - E(r_E)] \\
& \times \left[w_D^2 \sigma_D^2 + (1 - w_D)^2 \sigma_E^2 + 2w_D(1 - w_D)\text{Cov}(r_D, r_E) \right] \\
& = [E(r_D) - E(r_E)] \\
& \times [w_D^2 \sigma_D^2 + \sigma_E^2 - 2w_D \sigma_E^2 + w_D^2 \sigma_E^2 + 2w_D \text{Cov}(r_D, r_E) - 2w_D^2 \text{Cov}(r_D, r_E)] \\
& = [E(r_D) - E(r_E)] \times \{ w_D^2 [\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)] \\
& \quad + 2w_D [\text{Cov}(r_D, r_E) - \sigma_E^2] + \sigma_E^2 \} \\
& = [E(r_D) - E(r_E)] \times \{ w_D^2 [\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)] \} \\
& \quad + [E(r_D) - E(r_E)] \times \{ 2w_D [\text{Cov}(r_D, r_E) - \sigma_E^2] \} + [E(r_D) - E(r_E)] \times \sigma_E^2 \\
& = [E(r_D) - E(r_E)] \times [\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)] w_D^2 + 2[E(r_D) - E(r_E)] \\
& \quad \times [\text{Cov}(r_D, r_E) - \sigma_E^2] w_D + [E(r_D) - E(r_E)] \times \sigma_E^2
\end{aligned}$$

Right hand side of Equation A.33

$$\begin{aligned}
& [w_DE(r_D) + (1 - w_D)E(r_E) - r_f] \times [w_D \sigma_D^2 + w_D \sigma_E^2 \\
& - \sigma_E^2 + \text{Cov}(r_D, r_E) - 2w_D \text{Cov}(r_D, r_E)] \\
& = [w_DE(r_D) + E(r_E) - w_DE(r_E) - r_f] \times [w_D \sigma_D^2 \\
& \quad + w_D \sigma_E^2 - 2w_D \text{Cov}(r_D, r_E) - \sigma_E^2 + \text{Cov}(r_D, r_E)] \\
& = \{ w_D[E(r_D) - E(r_E)] + [E(r_E) - r_f] \} \times \{ w_D [\sigma_D^2 \\
& \quad + \sigma_E^2 - 2\text{Cov}(r_D, r_E)] + \text{Cov}(r_D, r_E) - \sigma_E^2 \} \\
& = w_D[E(r_D) - E(r_E)] \times w_D [\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)] \\
& \quad + w_D[E(r_D) - E(r_E)] \times [\text{Cov}(r_D, r_E) - \sigma_E^2] \\
& \quad + [E(r_E) - r_f] \times w_D [\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)] \\
& \quad + [E(r_E) - r_f] \times [\text{Cov}(r_D, r_E) - \sigma_E^2] \\
& = [E(r_D) - E(r_E)] \times [\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)] w_D^2 \\
& \quad + [E(r_D) - E(r_E)] \times [\text{Cov}(r_D, r_E) - \sigma_E^2] w_D \\
& \quad + [E(r_E) - r_f] \times [\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)] w_D \\
& \quad + [E(r_E) - r_f] \times [\text{Cov}(r_D, r_E) - \sigma_E^2]
\end{aligned}$$

Subtracting

$[E(r_D) - E(r_E)][\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)]w_D^2$ and $[E(r_D) - E(r_E)][\text{Cov}(r_D, r_E) - \sigma_E^2]w_D$ from both hand sides of Equation A.33, we have

$$\begin{aligned} & [E(r_D) - E(r_E)] \times [\text{Cov}(r_D, r_E) - \sigma_E^2]w_D + [E(r_D) - E(r_E)] \times \sigma_E^2 \\ &= [E(r_E) - r_f] \times [\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)]w_D + [E(r_E) - r_f] \\ & \quad \times [\text{Cov}(r_D, r_E) - \sigma_E^2] \end{aligned} \quad (\text{A.34})$$

Moving all the terms with w_D on one side and leaving the rest terms on the other side from Equation A.34, we have

$$\begin{aligned} & [E(r_D) - E(r_E)] \times \sigma_E^2 - [E(r_E) - r_f] \times [\text{Cov}(r_D, r_E) - \sigma_E^2] \\ &= [E(r_E) - r_f] \times [\sigma_D^2 + \sigma_E^2 - 2\text{Cov}(r_D, r_E)]w_D \\ & \quad - [E(r_D) - E(r_E)] \times [\text{Cov}(r_D, r_E) - \sigma_E^2]w_D \end{aligned} \quad (\text{A.35})$$

Rearrange Equation A.35 in order to solve for w_D , i.e.,

$$\begin{aligned} & [E(r_D) - E(r_E) + E(r_E) - r_f] \times \sigma_E^2 \\ & \quad - [E(r_E) - r_f] \text{Cov}(r_D, r_E) \\ &= \{ [E(r_E) - r_f] \sigma_D^2 + [E(r_E) - r_f] \sigma_E^2 \\ & \quad - [E(r_E) - r_f] [2\text{Cov}(r_D, r_E)] - [E(r_D) \\ & \quad - E(r_E)] \text{Cov}(r_D, r_E) + [E(r_D) - E(r_E)] \sigma_E^2 \} w_D \\ &= \{ [E(r_D) - r_f] \sigma_E^2 + [E(r_E) - r_f] \sigma_D^2 - [E(r_D) - r_f + E(r_E) - r_f] \text{Cov}(r_D, r_E) \} w_D \end{aligned}$$

Finally, we have the optimum weight of security D as

$$\begin{aligned} w_D &= \frac{[E(r_D) - r_f] \sigma_E^2 - [E(r_E) - r_f] \text{Cov}(r_D, r_E)}{[E(r_D) - r_f] \sigma_E^2 + [E(r_E) - r_f] \sigma_D^2} \\ & \quad - [E(r_D) - r_f + E(r_E) - r_f] \text{Cov}(r_D, r_E) \end{aligned}$$

Appendix G: Applications of the Binomial Distribution to Evaluate Call Options

In this appendix, we show how the binomial distribution is combined with some basic finance concepts to generate a model for determining the price of stock options.

G1. What is an Option?

In the most basic sense, an **option** is a contract conveying the right to buy or sell a designated security at a stipulated price. The contract normally expires at a pre-determined date. The most important aspect of an option contract is that the purchaser is under no obligation to buy; it is, indeed, an “option.” This attribute of an option contract distinguishes it from other financial contracts. For instance, whereas the holder of an option may let his or her claim expire unused if he or she so desires, other financial contracts (such as futures and forward contracts) obligate their parties to fulfill certain conditions.

A *call option* gives its owner the right to buy the underlying security, a *put option* the right to sell. The price at which the stock can be bought (for a call option) or sold (for a put option) is known as the exercise price.

G2. The Simple Binomial Option Pricing Model

Before discussing the binomial option model, we must recognize its two major underlying assumptions. First, the binomial approach assumes that trading takes place in discrete time, that is, on a period-by-period basis. Second, it is assumed that the stock price (the price of the underlying asset) can take on only two possible values each period; it can go up or go down.

Say we have a stock whose current price per share S can advance or decline during the next period by a factor of either u (up) or d (down). This price either will increase by the proportion $u - 1 \geq 0$ or will decrease by the proportion $1 - d$, $0 < d < 1$. Therefore, the value S in the next period will be either uS or dS . Next, suppose that a call option exists on this stock with a current price per share of C and an exercise price per share of X and that the option has one period left to maturity.

This option's value at expiration is determined by the price of its underlying stock and the exercise price X . The value is either

$$C_u = \text{Max}(0, uS - X) \quad (\text{A.36})$$

or

$$C_d = \text{Max}(0, dS - X) \quad (\text{A.37})$$

Why is the call worth $\text{Max}(0, uS - X)$ if the stock price is uS ? The option holder is not obliged to purchase the stock at the exercise price of X , so she or he will exercise the option only when it is beneficial to do so. This means the option can never have a negative value. When is it beneficial for the option holder to exercise the option? When the price per share of the stock is greater than the price per share at which he or she can purchase the stock by using the option, which is the exercise price, X . Thus if the stock price uS exceeds the exercise price X , the investor can exercise the option and buy the stock. Then he or she can immediately sell it for uS , making a profit of $uS - X$ (ignoring commission). Likewise, if the stock price declines to dS , the call is worth $\text{Max}(0, dS - X)$.

Also for the moment, we will assume that the risk-free interest rate for both borrowing and lending is equal to r percent over the one time period and that the exercise price of the option is equal to X .

To intuitively grasp the underlying concept of option pricing, we must set up a *risk-free portfolio* – a combination of assets that produces the same return in every state of the world over our chosen investment horizon. The investment horizon is assumed to be one period (the duration of this period can be any length of time, such as an hour, a day, a week, etc.). To do this, we buy h share of the stock and sell the call option at its current price of C . Moreover, we choose the value of h such that our portfolio will yield the same payoff whether the stock goes up or down.

$$h(uS) - C_u = h(dS) - C_d \quad (\text{A.38})$$

By solving for h , we can obtain the number of shares of stock we should buy for each call option we sell.

$$h = \frac{C_u - C_d}{(u - d)S} \quad (\text{A.39})$$

Here h is called the *hedge ratio*. Because our portfolio yields the same return under either of the two possible states for the stock, it is without risk and therefore should yield the risk-free rate of return, r percent, which is equal to the risk-free borrowing and lending rate, the condition must be true; otherwise, it would be possible to earn a risk-free profit without using any money. Therefore, the ending portfolio value must be equal to $(1 + r)$ times the beginning portfolio value, $hS - C$.

$$(1 + r)(hS - C) = h(uS) - C_u = h(dS) - C_d \quad (\text{A.40})$$

Note that S and C represent the beginning values of the stock price and the option price, respectively.

Setting $R = 1 + r$, rearranging to solve for C , and using the value of h from Equation A.39, we get

$$C = \left[\left(\frac{R - d}{u - d} \right) C_u + \left(\frac{u - R}{u - d} \right) C_d \right] / R \quad (\text{A.41})$$

where $d < r < u$. To simplify this equation, we set

$$p = \frac{R - d}{u - d} \text{ so } 1 - p = \left\{ \frac{u - R}{u - d} \right\} \quad (\text{A.42})$$

Thus we get the option's value with one period to expiration

$$C = \frac{pC_u + (1 - p)C_d}{R} \quad (\text{A.43})$$

This is the binomial call option valuation formula in its most basic form. In other words, this is the binomial valuation formula with one period to expiration of the option.

To illustrate the model's qualities, let's plug in the following values, while assuming the option has one period to expiration. Let

$$\begin{aligned} X &= \$100 \\ S &= \$100 \\ U &= (1.10), \text{ so } uS = \$110 \\ D &= (0.90), \text{ so } dS = \$90 \\ R &= 1 + r = 1 + 0.07 = 1.07 \end{aligned}$$

First we need to determine the two possible option values at maturity, as indicated in Table A.2.

Next we calculate the value of p as indicated in Equation A.42.

$$\begin{aligned} p &= \frac{1.07 - 0.90}{1.10 - 0.90} = 0.85 \text{ so } 1 - p = \frac{1.10 - 1.07}{1.10 - 0.90} \\ &= 0.15 \end{aligned}$$

Solving the binomial valuation equation as indicated in Equation A.43, we get

$$\begin{aligned} C &= \frac{0.85(10) + 0.15(0)}{1.07} \\ &= \$7.94 \end{aligned}$$

Table A.2 Possible option value at maturity

Today Stock (S)	Option (C)	Next Period (Maturity)
\$100 C	$uS = \$110$	$C_u = \text{Max}(0, uS - X)$ $= \text{Max}(0, 110 - 100)$ $= \text{Max}(0, 10)$ $= \$10$
	$dS = \$90$	$C_d = \text{Max}(0, dS - X)$ $= \text{Max}(0, 90 - 100)$ $= \text{Max}(0, -10)$ $= \$0$

The correct value for this particular call option today, under the specified conditions, is \$7.94. If the call option does not sell for \$7.94, it will be possible to earn arbitrage profits. That is, it will be possible for the investor to earn a risk-free profit while using none of his or her own money. Clearly, this type of opportunity cannot continue to exist indefinitely.

G3. The Generalized Binomial Option Pricing Model

Suppose we are interested in the case where there is more than one period until the option expires. We can extend the one-period binomial model to consideration of two or more periods.

Because we are assuming that the stock follows a binomial process, from one period to the next it can only go up by a factor of u or go down by a factor of d . After one period the stock's price is either uS or dS . Between the first and second periods, the stock's price can once again go up by u or down by d , so the possible prices for the stock two periods from now are uuS , udS , and ddS . This process is demonstrated in tree diagram (Figure A.1) given in Example A.1 later in this appendix.

Note that the option's price at expiration, two periods from now, is a function of the same relationship that determined its expiration price in the one-period model, more specifically, the call option's maturity value is always

$$C_T = [0, S_T - X]$$

(A.44)

where T designated the maturity date of the option.

To derive the option's price with two periods to go ($T = 2$), it is helpful as an intermediate step to derive the value of C_u and C_d with one period to expiration when the stock price is either uS or dS , respectively.

$$C_u = \frac{pC_{uu} + (1-p)C_{ud}}{R} \quad (\text{A.45})$$

$$C_d = \frac{pC_{du} + (1-p)C_{dd}}{R} \quad (\text{A.46})$$

Equation A.45 tells us that if the value of the option after one period is C_u , the option will be worth either C_{uu} (if the stock price goes up) or C_{ud} (if stock price goes down) after one more period (at its expiration date). Similarly, Equation A.46 shows that the value of the option is C_d after one period, the option will be worth either C_{du} or C_{dd} at the end of the second period. Replacing C_u and C_d in Equation A.43 with their expressions in Equations A.45 and A.46, respectively, we can simplify the resulting equation to yield the two-period equivalent of the one-period binomial pricing formula, which is

$$C = \frac{p^2C_{uu} + 2p(1-p)C_{ud} + (1-p)^2C_{dd}}{R^2} \quad (\text{A.47})$$

In Equation A.47, we used the fact that $C_{ud} = C_{du}$ because the price will be the same in either case.

We know the values of the parameters S and X . If we assume that R , u , and d will remain constant over time, the possible maturity values for the option can be determined exactly. Thus deriving the option's fair value with two periods to maturity is a relatively simple process of working backwards from the possible maturity values.

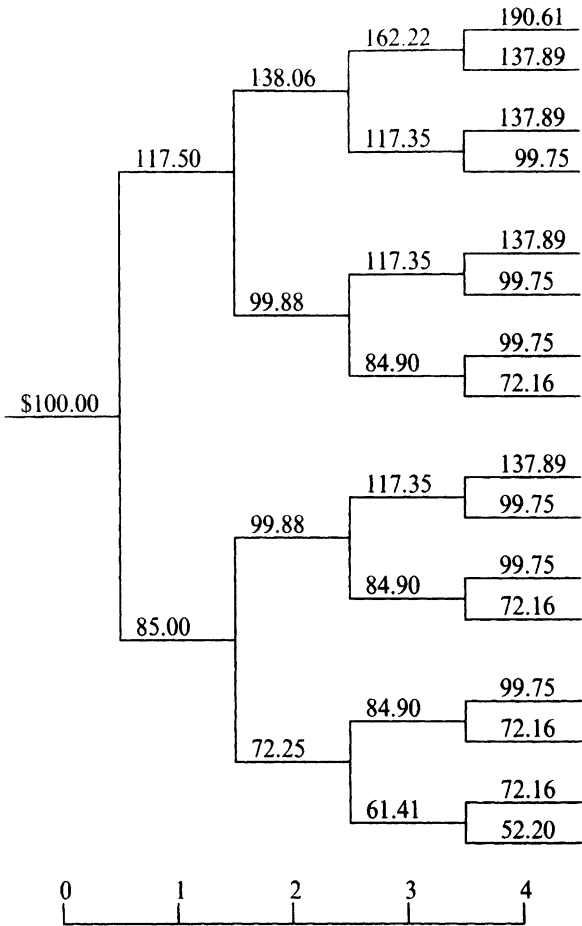
Using this same procedure of going from a one-period model to a two-period model, we can extend the binomial approach to its more generalized form, with n periods maturity

$$C = \frac{1}{R^n} \sum_{k=0}^n \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \text{Max}[0, u^k d^{n-k} S - X] \quad (\text{A.48})$$

To actually get this form of the binomial model, we could extend the two-period model to three periods, then from three periods to four periods, and so on. Equation A.48 would be the result of these efforts. To show how Equation A.48 can be used to assess a call option's value, we modify the example as follows: $S = \$100$, $X = \$100$, $R = 1.07$, $n = 3$, $u = 1.1$ and $d = 0.90$.

First we calculate the value of p from Equation A.42 as 0.85, so $1 - p$ is 0.15. Next we calculate the four possible ending values for the call option after three periods in terms of $\text{Max}[0, u^k d^{n-k} S - X]$.

Fig. A.1 Price path of underlying stock (Source: Rendelman and Bartter, 1979, 1906)



$$C_1 = \left[0, (1.1)^3 (0.90)^0 (100) - 100 \right] = 33.10$$
$$C_2 = \left[0, (1.1)^2 (0.90) (100) - 100 \right] = 8.90$$
$$C_3 = \left[0, (1.1) (0.90)^2 (100) - 100 \right] = 0$$
$$C_4 = \left[0, (1.1)^0 (0.90)^3 (100) - 100 \right] = 0$$

Now we insert these numbers (C_1 , C_2 , C_3 , and C_4) into the model and sum the terms.

$$\begin{aligned}
C &= \frac{1}{(1.07)^3} \left[\frac{3!}{0!3!} (0.85)^0 (0.15)^3 \times 0 \right. \\
&\quad + \frac{3!}{1!2!} (0.85)^1 (0.15)^2 \times 0 \\
&\quad + \frac{3!}{2!1!} (0.85)^2 (0.15)^1 \times 8.90 \\
&\quad \left. + \frac{3!}{3!0!} (0.85)^3 (0.15)^0 \times 33.10 \right] \\
&= \frac{1}{1.225} \left[0 + 0 + \frac{3 \times 2 \times 1}{2 \times 1 \times 1} (0.7225)(0.15)(8.90) \right. \\
&\quad \left. + \frac{3 \times 2 \times 1}{3 \times 2 \times 1 \times 1} \times (0.61413)(1)(33.10) \right] \\
&= \frac{1}{1.225} [(0.32513 \times 8.90) + (0.61413 \times 33.10)] \\
&= \$18.96
\end{aligned}$$

As this example suggests, working out a multiple-period problem by hand with this formula can become laborious as the number of periods increases. Fortunately, programming this model into a computer is not too difficult.

Now let's derive a binomial option pricing model in terms of the cumulative binomial density function. As a first step, we can rewrite Equation A.48 as

$$\begin{aligned}
C &= S \left[\sum_{k=m}^n \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{u^k d^{n-k}}{R^n} \right] \\
&\quad - \frac{X}{R^n} \left[\sum_{k=m}^n \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \right]
\end{aligned} \tag{A.49}$$

This formula is identical to Equation A.48 except that we have removed the Max operator. In order to remove the Max operator, we need to make $u^k d^{n-k} S - X$ positive, which we can do by changing the counter in the summation from $k = 0$ to $k = m$. What is m ? It is the minimum number of upward stock movements necessary for the option to terminate "in the money" (that is, $u^k d^{n-k} S - X > 0$). How can we interpret Equation A.49? Consider the second term in brackets; it is just a cumulative binomial distribution with parameters of n and p . Likewise, via a small algebraic manipulation we can show that the first term in the brackets is also a cumulative binomial distribution. This can be done by defining $P' \equiv (u/R)p$ and $1 - P' \equiv (d/R)(1 - p)$. Thus

$$p^k (1-p)^{n-k} \frac{u^k d^{n-k}}{R^n} = p'^k (1-p')^{n-k}$$

Therefore the first term in brackets is also a cumulative binomial distribution with parameters of n and p' . Using Equation A.45 in the text, we can write the binomial call option model as

$$C = SB_1(n, p', m) - \frac{X}{R^n} B_2(n, p, m) \quad (\text{A.50})$$

where

$$B_1(n, p', m) = \sum_{k=m}^n C_k^n p'^k (1 - p')^{n-k}$$

$$B_2(n, p, m) = \sum_{k=m}^n C_k^n p^k (1 - p)^{n-k}$$

and m is the minimum amount of time the stock has to go up for the investor to finish *in the money* (that is, for the stock price to become larger than the exercise price).

In this appendix, we showed that by employing the definition of a call option and by making some simplifying assumptions, we could use the binomial distribution to find the value of a call option. In the next chapter, we will show how the binomial distribution is related to the normal distribution and how this relationship can be used to derive one of the most famous valuation equations in finance, the Black-Scholes option pricing model.

Example A.1

A Decision Tree Approach to Analyzing Future Stock Price

By making some simplifying assumptions about how a stock's price can change from one period to the next, it is possible to forecast the future price of the stock by means of a decision tree. To illustrate this point, let's consider the following example.

Suppose the price of Company A's stock is currently \$100. Now let's assume that from one period to the next, the stock can go up by 17.5% or go down by 15%. In addition, let us assume that there is a 50% chance that the stock will go up and a 50% chance that the stock will go down. It is also assumed that the price movement of a stock (or of the stock market) today is completely independent of its movement in the past; in other words, the price will rise or fall today by a random amount. A sequence of these random increases and decreases is known as a **random walk**.

Given this information, we can lay out the paths that the stock's price may take. Figure A.1 shows the possible stock prices for company A for four periods.

Note that in period 1 there are two possible outcomes: the stock can go up in value by 17.5% to \$117.50 or down by 15% to \$85.00. In period 2 there are four possible outcomes. If the stock went up in the first period, it can go up again to \$138.06 or down in the second period to \$99.88. Likewise, if the stock went down in the first period, it can go down again to \$72.25 or up in the second period to \$99.88. Using the same argument, we can trace the path of the stock's price for all four periods.

If we are interested in forecasting the stock's price at the end of period 4, we can find the average price of the stock for the 16 possible outcomes that can occur in period 4.

$$\bar{P} = \frac{\sum_{i=1}^{16} P_i}{16} = \frac{190.61 + 137.89 + \cdots + 52.20}{16} = 105.09$$

We can also find the standard deviation for the stock's return.

$$\begin{aligned}\sigma_P &= \left[\frac{(190.61 - 105.09)^2 + \cdots + (52.20 - 105.09)^2}{16} \right]^{1/2} \\ &= 34.39\end{aligned}$$

$\bar{P}$ and σ_P can be used to predict the future price of stock A.

Appendix H: Derivation of Modigliani and Miller (M&M) Proposition I and II with Taxes

H1. M&M Proposition I with Taxes

Assume that the firms are non-growth companies; the market value of levered firm is equal to the market value of unlevered firm plus the present value of the cost of perpetually total debt.

$$V_L = V_U + \sum_{t=1}^{\infty} \frac{TDk_d}{(1+k_d)^t} = V_U + \frac{TDk_d}{k_d} = V_U + TD \quad (\text{A.51})$$

Where V_L = market value of levered firm, V_U = market value of unlevered firm, T = marginal corporate tax rate, D = total debt, k_d = the cost of debt, and TD = tax shield value.

Alternatively, the market value of levered firm, V_L , can be viewed as the total cash flow to all stakeholders

$$\begin{aligned} & (EBIT - k_d D) \times (1 - T) + k_d D \\ & = EBIT \times (1 - T) + T k_d D \end{aligned} \quad (\text{A.52})$$

The cash flow to all stakeholders is made up of cash flow to stockholders plus cash flow to bondholders. The present value of first term is equal to the market value of unlevered firm, V_U , and the present value of second term is TD . Therefore, $V_L = V_U + TD$.

Miller (1977) modified Eq. A.51 by introducing personal as well as corporate taxes into the model, and obtaining

$$V_L = V_U + \left[1 - \left[\frac{(1-T)(1-T^{PS})}{(1-T^{PD})} \right] \right] D \quad (\text{A.53})$$

Where V_L , V_U , T , and D have been defined in Eq. A.51; T^{PS} and T^{PD} are the personal tax rate on equity income and the personal tax from bond income, respectively. If T^{PS} equals to T^{PD} , then Eq. A.53 can be reduced to M&M Proposition I with taxes.

H2. M&M Proposition II with Taxes

Since market value of levered firm, V_L , is equal to total equity, E , plus total debt, D , and based on M&M proposition I with taxes, the market value of unlevered firm can be derived as

$$V_L = E + D = V_U + TD \Rightarrow V_U = E + (1 - T)D \quad (\text{A.54})$$

The cash flow from each side of balance sheet must equal, therefore

$$\begin{aligned} Ek_e + Dk_d &= V_U k_u + TDk_d \\ &= (E + (1 - T)D)k_u + TDk_d \end{aligned} \quad (\text{A.55})$$

Where E = total equity, D = total debt, k_e = the cost of equity, k_d = the cost of debt, k_u = the cost of unlevered equity, and T = marginal corporate tax rate.

Divide both side by total equity, E , then we can get

$$\begin{aligned} k_e + \frac{D}{E}k_d &= \left(1 + (1 - T)\frac{D}{E}\right)k_u + \frac{D}{E}Tk_d \\ &= k_u + (1 - T)\frac{D}{E}k_u + \frac{D}{E}Tk_d \end{aligned} \quad (\text{A.56})$$

$$\begin{aligned} k_e &= k_u + (1 - T)\frac{D}{E}k_u + \frac{D}{E}Tk_d - \frac{D}{E}k_d \\ &= k_u + (1 - T)\frac{D}{E}k_u - (1 - T)\frac{D}{E}k_d \\ &= k_u + (k_u - k_d)\frac{D}{E}(1 - T) \end{aligned} \quad (\text{A.57})$$

Figure A.2 represents the relationship between the cost of equity with and without taxes, and the cost of unlevered firm. The weighted average cost of capital with taxes will decrease when the ratio of debt to equity increases.

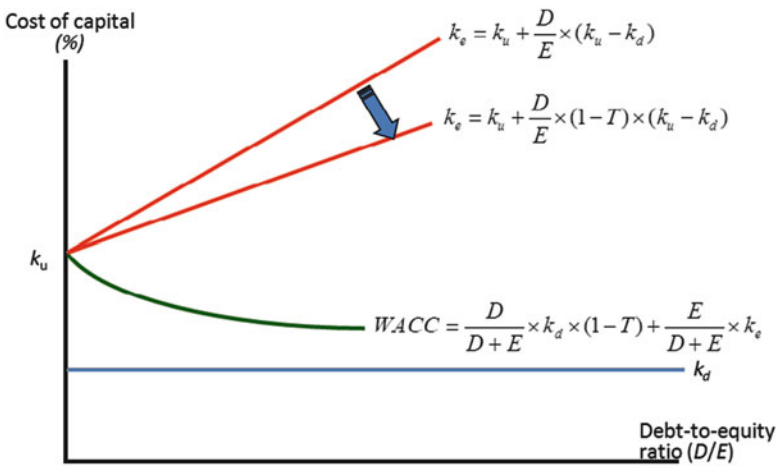


Fig. A.2 The Relationship Between the Cost of Capital and Debt-to-Equity Ratio

Appendix I: Derivation of Capital Market Line (CML)

By geometric theory, triangles R_fPE and R_fM_pD in the Fig. A.3 below are similar and are, therefore, directly proportional,

$$\Delta PR_fE \cong \Delta M_pR_fD \quad (\text{A.58})$$

Therefore,

$$\begin{aligned} \frac{E(R_p) - R_f}{E(R_{M_p}) - R_f} &= \frac{\sigma_p}{\sigma_{M_p}} \\ \Rightarrow (E(R_p) - R_f) &= \frac{\sigma_p}{\sigma_{M_p}} (E(R_{M_p}) - R_f) \\ \Rightarrow E(R_p) &= R_f + [E(R_{M_p}) - R_f] \frac{\sigma_p}{\sigma_{M_p}} \end{aligned} \quad (\text{A.59})$$

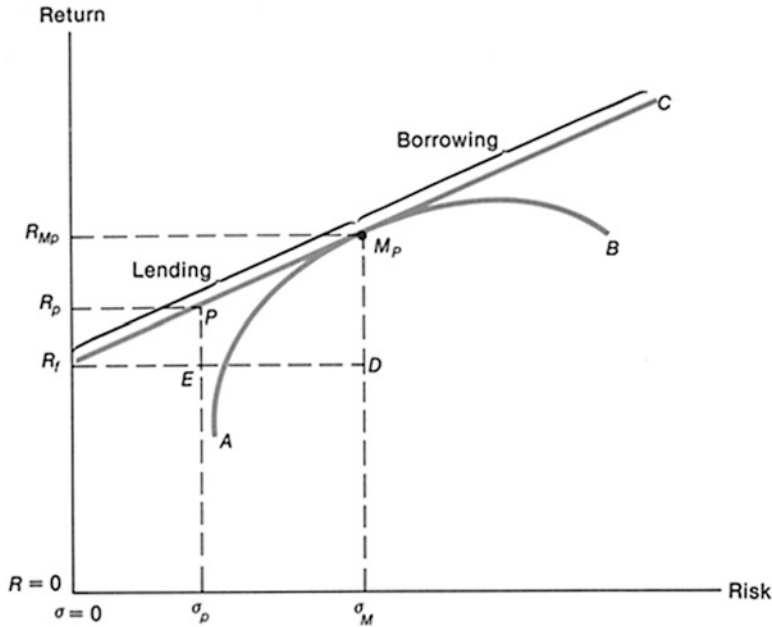


Fig. A.3 The Capital Market Line

where R_f = the risk-free rate; R_{M_p} = return on market portfolio M_p ; R_p = return on the portfolio consisting of combinations of the risk-free asset and portfolio M_p ; σ_p and σ_{M_p} = standard deviations of the portfolio and the market; and the operator E denotes expectations.

Appendix J: Derivation of Capital Market Line (SML)

Sharpe (1964) used a general risky asset that did not lie on the CML and dubbed it *I* in Fig. A.4. The combinations of risk and return possible by combining security *I* with the market portfolio, *M*, are shown by figure above. The average return and standard deviation for any *I-M* combination can be approached in the same way as for a two-asset case:

$$E(R_P) = w_i E(R_i) + (1 - w_i) E(R_m) \quad (\text{A.60})$$

$$\sigma_P = [w_i^2 \sigma_i^2 + (1 - w_i)^2 \sigma_m^2 + 2(1 - w_i) w_i \sigma_{im}]^{\frac{1}{2}} \quad (\text{A.61})$$

where w_i represents excess demand for *I* or demand greater than its equilibrium weight in portfolio *M*, and σ_{im} is the covariance of *i* and *m*.

The change in mean and standard deviation as the proportion w_i changes are the partial derivatives

$$\frac{\partial E(R_P)}{\partial w_i} = E(R_i) - E(R_m) \quad (\text{A.62})$$

$$\frac{\partial(\sigma_P)}{\partial w_i} = \frac{1}{2} [w_i^2 \sigma_i^2 + (1 - w_i)^2 \sigma_m^2 + 2w_i(1 - w_i) \sigma_{im}]^{-\frac{1}{2}} (2w_i \sigma_i^2 - 2\sigma_m^2 + 2w_i \sigma_m^2 + 2\sigma_{im} - 4w_i \sigma_{im}) \quad (\text{A.63})$$

When $w_i = 0$, the security is held in proportion to its total market value and there is no excess demand for security *I*. This is the key insight to Sharpe's paper, for when $w_i = 0$, it is possible to equate the slope of the curve *IMI* ' with the capital market line and thus obtain an expression for the return on any risky security *I*. At equilibrium when $w_i = 0$, the slope along the *IMI* ' curve will equal

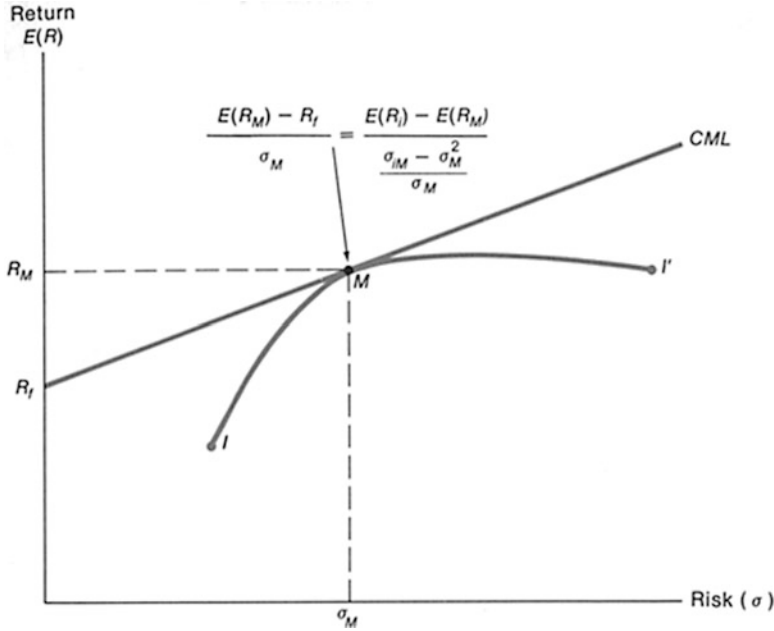


Fig. A.4 The Opportunity Set Provided by Combinations of Risky Asset (I) and Market Portfolio (M)

$$\frac{\partial E(R_P)}{\partial(\sigma_P)} = \frac{\frac{\partial E(R_P)}{\partial w_i}}{\frac{\partial(\sigma_P)}{\partial w_i}} = \frac{\frac{E(R_i) - E(R_M)}{\sigma_{im} - \sigma_m^2}}{\sigma_m} \quad (\text{A.64})$$

The slope of the capital market line at point M is

$$\frac{E(R_m) - R_f}{\sigma_m} \quad (\text{A.65})$$

Let Eq. A.64 equal to Eq. A.65, then rearranging the terms to solve for $E(R_i)$ gives the equation for the security market line or CAPM

$$E(R_i) = R_f + [E(R_m) - R_f] \frac{\sigma_{im}}{\sigma_m^2} \quad (\text{A.66})$$

This represents the return on any risky asset I . At equilibrium, every risky asset will be priced so that it lies along the security market line. It should be noted that the term σ_{im}/σ_m^2 represents the beta coefficient for the regression of R_i vs. R_m , so that Eq. A.66 can be rewritten as

$$E(R_i) = R_f + [E(R_m) - R_f] \beta_i \quad (\text{A.67})$$

Appendix K: Derivation of Black-Scholes Option Pricing Model

Assume the stock prices follow a lognormal distribution and denote the current stock price by S and the stock price at the end of t -th period by S_t then $\frac{S_t}{S_{t-1}} = \exp(K_t)$ is a random variable with a lognormal distribution, where K_t is the rate of return in t -th period and follows normal distribution with the constant mean μ and variance σ^2 , therefore,

$$\begin{aligned} E\left[\frac{S_T}{S}\right] &= E\left[\frac{S_1}{S} \frac{S_2}{S_1} \cdots \frac{S_T}{S_{T-1}}\right] \\ &= E[\exp(K_1 + K_2 + \cdots + K_T)] = T\mu + \frac{T\sigma^2}{2} \end{aligned} \quad (\text{A.68})$$

Under the assumption of a risk-neutral investor, the expected return $E\left[\frac{S_T}{S}\right]$ is assumed to be $\exp[rT]$ (where r is the riskless rate of interest). In other words, $\mu = r - \sigma^2/2$.

The call option price C can be determined by discounting the expected value of the terminal option price by the riskless rate of interest (r):

$$C = \exp[-rT]E[\text{Max}(S_T - X, 0)] \quad (\text{A.69})$$

where T is the time of expiration and X is the striking price, r is the riskless interest rate, S_T is the stock price at time T . Note that

$$\text{Max}(S_T - X, 0) = \left(S\left(\frac{S_T}{S} - \frac{X}{S}\right)\right) \quad \text{for } \frac{S_T}{S} > \frac{X}{S} \quad (\text{A.70})$$

Let $y = \frac{S_T}{S}$ has a lognormal distribution with mean $\mu T = (r - \frac{1}{2}\sigma^2)T$ and variance $\sigma^2 T$, then

$$\begin{aligned}
C &= \exp[-rT]E[\text{Max}(S_T - X, 0)] \\
&= \exp[-rT] \int_{\frac{X}{S}}^{\infty} S \left[y - \frac{X}{S} \right] g(y) dy \\
&= S \exp[-rT] \int_{\frac{X}{S}}^{\infty} y g(y) dy - \exp[-rT] S \frac{X}{S} \int_{\frac{X}{S}}^{\infty} g(y) dy
\end{aligned} \tag{A.71}$$

Let $x = \ln(y)$, then x follows normal distribution with mean $\mu T = (r - \frac{1}{2}\sigma^2)T$ and variance $\sigma^2 T$, and

$$\begin{aligned}
dx &= \frac{1}{y} dy, \frac{f(x)}{y} = g(y) \\
\int_{\frac{X}{S}}^{\infty} g(y) dy &= \int_{\ln(\frac{X}{S})}^{\infty} \frac{f(x)}{y} (y dx) \\
&= \int_{\ln(\frac{X}{S})}^{\infty} f(x) dx \\
&= \int_{\frac{\ln(\frac{X}{S}) - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}}^{\infty} h(z) dz
\end{aligned} \tag{A.72}$$

Where $g(y)$ is the probability density function of y , $f(x)$ is the probability density function of x , z is standard normal distribution, and $h(z)$ is the probability density function of z . The first term of call option can be derived as

$$\begin{aligned}
\int_{\frac{X}{S}}^{\infty} e^{-rT} y g(y) dy &= \int_{\ln(\frac{X}{S})}^{\infty} e^{-rT} e^x \frac{f(x)}{y} (y dx) \\
&= \int_{\ln(\frac{X}{S})}^{\infty} e^{-rT} e^x f(x) dx \\
&= \int_{\ln(\frac{X}{S})}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 T}} e^{\left(\frac{-(x - (r - \frac{1}{2}\sigma^2)T)^2}{2\sigma^2 T} - rT + x \right)} dx \\
&= \int_{\ln(\frac{X}{S})}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 T}} e^{\frac{-(x - (r - \frac{1}{2}\sigma^2)T)^2}{2\sigma^2 T}} dx \\
&= \int_{\frac{\ln(\frac{X}{S}) - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}}^{\infty} h(z) dz
\end{aligned} \tag{A.73}$$

Where z is standard normal distribution, and $h(z)$ is the probability density function of z . Therefore, the call option pricing model can be rewritten as

$$\begin{aligned}
 C &= S \int_{\frac{\ln\left(\frac{X}{S}\right) - \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}}^{\infty} h(z) dz \\
 &\quad - Xe^{-rT} \int_{\frac{\ln\left(\frac{X}{S}\right) - \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}}^{\infty} h(z) dz \\
 &= SN(d_1) - Xe^{-rT}N(d_2)
 \end{aligned} \tag{A.74}$$

where

$$\begin{aligned}
 d_1 &= - \left[\frac{\ln\left(\frac{X}{S}\right) - \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} \right] = \frac{\ln\left(\frac{S}{X}\right) + \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}, \\
 d_2 &= - \left[\frac{\ln\left(\frac{X}{S}\right) - \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} \right] = \frac{\ln\left(\frac{S}{X}\right) + \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}
 \end{aligned}$$

Based on put-call parity, it can be shown that the relationship between a call option (C) and a put option (P) can be defined as

$$C + Xe^{-rT} = P + S \tag{A.75}$$

Substituting Eqs. A.74 into A.75, we obtain the put option formula as

$$P = Xe^{-rT}N(-d_2) - SN(-d_1) \tag{A.76}$$

where S , C , r , T , d_1 and d_2 are identical to those defined in the call option model.

Appendix L: Hillier's Statistical Distribution Method

The riskiness of an investment proposal is generally defined as the variability of its possible returns. Decision situations can generally be classified into three types: certainty, risk, and uncertainty. The distinction between risk and uncertainty is that risk involves situations in which the probabilities of a particular event occurring are known, whereas with uncertainty these probabilities themselves are not known. Throughout this chapter, the terms “risk” and “uncertainty” will be used interchangeably, although the strict difference between them is not a matter to be taken lightly.

Economic factors peculiar to investment, competition, technological development, consumer preferences, and labor market conditions are a few of the factors that make it virtually impossible to foretell the future. Consequently, the revenues, costs, and economic life of investment projects are less than certain.

With the introduction of risk, a firm is no longer indifferent between two investment proposals having equal net present values or internal rates of return. Both net present value and its standard deviation (σ_{NPV}) should be estimated in performing capital-budgeting analyses under uncertainty. A NPV approach under uncertainty can be defined as

$$\tilde{NPV} = \sum_{t=1}^N \frac{\bar{C}_t}{(1+k)^t} + \frac{S_n}{(1+k)^N} - I_0 \quad (\text{A.77})$$

where C_t is the uncertain net cash flow in period t , k is the risk-adjusted discount rate, S_n is the salvage value, and I_0 is the initial outlay. The mean of the NPV distribution and its standard deviation can be defined as:

$$\overline{NPV} = \sum_{t=1}^N \frac{\bar{C}_t}{(1+k)^t} + \frac{S_n}{(1+k)^N} - I_0 \quad (\text{A.78})$$

$$\sigma_{NPV} = \left[\sum_{t=1}^N \frac{\sigma_t^2}{(1+k)^{2t}} + \sum_{\tau=1}^N \sum_{t=1}^N w_t w_\tau \text{Cov}(C_\tau, C_t) \right]^{1/2} \quad (\tau \neq t), \quad (\text{A.79})$$

where σ is the variance of the cash flow in the t th period and w_t and w_τ are the discount factors in the t th and τ th periods, respectively. The calculation of σ_{NPV} is similar to the standard deviation of a portfolio's returns.

$\text{Cov}(C_\tau, C_t)$ is used to measure the covariability between the cash flow in the t th and τ th periods. It is well known that:

$$\text{Cov}(C_t, C_\tau) = \rho_{\tau t} \sigma_\tau \sigma_t, \quad (\text{A.80})$$

where $\rho_{\tau t}$ is the correlation coefficient between the cash flow in the t th period and that in the τ th period. In one special case, the cash flows are perfectly correlated over time. If such is the case, then the deviation of the actual cash flow from the mean of its corresponding profitability distribution of expected cash flows for the period gives us the information that cash flows in future periods will deviate in a similar fashion. In other words, Eq. (A.79) can be rewritten as:

$$\sigma_{\text{NPV}} = \sum_{t=1}^N \frac{\sigma_t}{(1+k)^t} \quad (\text{A.81})$$

If the cash flows are mutually independent, then Eq. (A.81) reduces to:

$$\sigma_{\text{NPV}} = \sqrt{\sum_{t=1}^N \frac{\sigma_t^2}{(1+k)^{2t}}} \quad (\text{A.82})$$

Hillier (1963) combines the assumption of mutual independence and perfect correlation in developing a model to deal with mixed situations. The model as defined in Eq. (A.83) can be used to analyze investment proposals in which some of the expected cash flows are closely related, and others are fairly independent.

$$\sigma = \sqrt{\sum_{t=1}^N \frac{\sigma_{yt}^2}{(1+k)^{2t}} + \sum_h^m \left(\sum_{t=1}^N \frac{\sigma_{zt}^h}{(1+k)^t} \right)^2} \quad (\text{A.83})$$

where σ_{yt}^2 is the variance for an independent net cash flow in period t , and σ_{zt}^h is the standard deviation for stream h of a perfectly correlated stream of cash flows t . It is clear that Eq. (A.83) is a combination of Eqs. (A.81) and (A.82). Chen and Moore (1982) have generalized this model by introducing the estimation risk.

They argue that Hillier's (1963) approach is unrealistic due to the assumption of parameter certainty. In reality, the parameters of the probability distribution are never known with certainty, and thus the capital budgeting decisions based on Hillier's approach may be potentially misleading because this approach ignores estimation risk. Chen and Moore (1982) hence extend Hillier's approach with a Bayesian framework which leads to the derivation of a predictive distribution that is not conditioned on unknown parameters. Therefore, the effect of estimation risk in capital budgeting decisions can be eliminated in Chen and Moore (1982) approach.

In addition, Wang and Lee (2010) provide the fuzzy real option valuation approach to solve the capital budgeting decisions under uncertainty environment. In Wang and Lee's model framework, the concept of probability is employed in describing fuzzy events and the estimated cash flows are based on the fuzzy numbers which can better reflect the uncertainty in the project. By using a fuzzy real option valuation, the managers can select fuzzy projects and determine the optimal time to abandon the project under the assumption of limited capital budget.

To illustrate the model of (A.83), suppose that a firm is considering the introduction of a new product with returns expected over the next 5 years. Because the product's market reception is uncertain, management feels that, if initial reception exceeds expectations, actual acceptance in later years will also exceed expectations, in approximately the same proportions. For simplicity, it is believed that the net marketing cash flow (sales less marketing and advertising expenses) can be treated as perfectly correlated over time.

On the other hand, estimates of the initial investment in the project and the production costs are reasonably reliable, so that any deviation from expectations is assumed to be attributable to random fluctuations. Consequently, initial investment and net production cash flows are regarded as being mutually independent over time. The probability information for the introduction of the new product is shown in Table A.3. We assume that each of the probability distributions involved can be regarded as normal. If 10% is used as the risk-free rate, the expected value of the net present value for the proposal is:

$$\begin{aligned} NPV &= \sum_{t=0}^5 \frac{C_t}{(1.04)^t} = -600 + \frac{300 - 250}{(1.04)} + \frac{600 - 200}{(1.04)^2} + \frac{500 - 200}{(1.04)^3} \\ &\quad + \frac{400 - 200}{(1.04)^4} + \frac{300 - 100}{(1.04)^5} \\ &= \$419.95 = \$420.00. \end{aligned}$$

Table A.3 Expected cash flow for new product

Year	Source	Expected value of net cash flow (in thousands)	Standard deviation (in thousands)
0	Initial investment	−\$600	\$50
1	Production cash outflow	−250	20
2	Production cash outflow	−200	10
3	Production cash outflow	−200	10
4	Production cash outflow	−200	10
5	Production outflow – salvage value	−100	15
1	Marketing	300	50
2	Marketing	600	100
3	Marketing	500	100
4	Marketing	400	100
5	Marketing	300	100

Following Eq. (A.79), we can calculate the standard deviation as:

$$\begin{aligned}\sigma_{\text{NPV}} &= \left[50^2 + \frac{20^2}{(1.04)^2} + \dots + \frac{15^2}{(1.04)^{10}} + \left(\frac{50}{1.04} + \dots + \frac{100}{(1.04)^5} \right)^2 \right]^{1/2} \\ &= \$398.\end{aligned}$$

Thus, the expected net present value of the proposal is \$420 and the standard deviation of the distribution relating to NPV is \$398. Using this information, we can express an interval inference for NPV as:

$$\Pr(\$420 \pm \$398) = 68.27\%;$$

$$\Pr(\$420 \pm (2)(\$398)) = 95.45\%;$$

$$\Pr(\$420 \pm (3)(\$398)) = 99.73\%.$$

Hillier's approach takes us a long way toward coping with the correlation of cash flows over time, but his method is not always practically applicable. For many investment proposals, cash flows fall into neither of these categories, since in practice they show less than perfect correlation over time. If the correlation is moderate, then the classification suggested by Hillier will not be appropriate. The method suggested by Bonini (1975) and others for handling the problem of moderate correlation is with a series of conditional-probability distributions. Wang and Lee (2010) have expended Bonini's result in terms of fuzzy set and real option approach.

To illustrate the use of this model, we refer to Table A.4, where we consider a project requiring an initial outlay of \$10,000 (Column 1). In each of the following time periods (Columns 3, 5, and 7), the cash flow to be received is not known with perfect certainty, but the probabilities associated with each cash flow in each period are assumed to be known (Columns 2, 4, and 6), so that we are dealing with a case of risk, and not strict uncertainty. We note that Columns (4) and (6) are conditional-probability figures, where the later periods' expected cash flows depend highly on what occurs in earlier time periods. Given our cash flow and simple probability estimates for the periods 1, 2, and 3, we find 27 possible joint probabilities (Column 8), each of which corresponds to a cash-flow series. Thus in the uppermost path, we find the joint probability 0.015 being associated with a cash flow of \$2000 in Period 1, followed by cash flows of \$2000 and \$1000 in Periods 2 and 3, respectively. This particular path is the worst possible result; in nondiscounted dollar terms, there is a 50% loss on the investment. At the other end of the spectrum, we find the best possible outcome offers a return \$12,945 (the last line of Table A.5) with a joint probability of 0.0125. The returns and joint probabilities can be calculated similarly for the other 25 possible patterns.

While the primary hurdle in this process is in estimating the cash flows and the probabilities, we recognize that it is the NPV figures we are ultimately interested

Table A.4 Illustration of conditional-probability distribution approach

Initial outlay period 0	Initial probability $P(1)$	Net cash flow	Conditional probability $P(3 2,1)$	Net cash flow	Conditional probability $P(3 2,1)$	Cash flow	Joint probability
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					0.2	1000	0.015
			0.25	2000	0.5	2000	0.0375
					0.3	3000	0.0225
					0.2	2000	0.03
10,000	0.3	2000	0.5	3000	0.5	3000	0.075
					0.3	4000	0.045
					0.2	3000	0.015
			0.25	4000	0.5	4000	0.0375
					0.3	5000	0.0225
					0.3	2000	0.045
			0.3	3000	0.4	4000	0.06
					0.3	6000	0.045
					0.3	4000	0.06
	0.5	4000	0.4	5000	0.4	5000	0.08
					0.3	6000	0.06
					0.3	5000	0.045
			0.3	7000	0.4	7000	0.06
					0.3	9000	0.045
					0.25	4000	0.0125
			0.25	5000	0.5	6000	0.025
					0.25	8000	0.0125
					0.25	5000	0.025
	0.2	6000	0.5	7000	0.5	7000	0.05
					0.25	9000	0.025
					0.25	7000	0.0125
			0.25	8000	0.5	9000	0.025
					0.25	11,000	0.0125

in. Table A.5 gives the present value of the cash flows and the net present value of the project for each cash-flow series, where a constant 4% discount rate was employed. When we multiply each joint probability by the expected net present value associated with that probability, we obtain the expected net present value for the project as a whole, there shown to be \$2517.18. Also, from all 27 expected net-present-value figures, we can calculate the variance and the standard deviation for the project as a whole, shown in the lower portion of Table A.5 as \$20,359,090.91 and \$4512.11, respectively, where these figures can be used to assess the riskiness of the project.

The use of this form of conditional-probability distribution enables us to take account of the dependence or correlation in the cash flows over time. In the above

Table A.5 NPV and joint probability

PVA	NPV	Probability
4661.2	−5338.8	0.015
5550.2	−4449.8	0.0375
6439.2	−3560.8	0.0225
6474.76	−3525.24	0.03
7363.76	−2636.24	0.075
8252.76	−1747.24	0.045
8288.32	−1711.68	0.015
9177.32	−822.68	0.0375
10,066.32	66.32	0.0225
8297.84	−1602.16	0.045
10,175.84	175.84	0.06
11,953.84	1953.84	0.045
12,024.96	2024.96	0.06
12,913.96	2913.96	0.08
13,802.96	3802.96	0.06
14,763.08	4763.08	0.045
16,541.08	6541.08	0.06
18,319.08	8319.08	0.045
13,948.04	3948.04	0.0125
15,726.04	5726.04	0.025
17,504.04	7504.04	0.0125
16,686.16	6686.16	0.025
18,464.16	8464.16	0.05
20242.16	10,242.16	0.025
19,388.72	9388.72	0.0125
21,166.72	11,166.72	0.025
22,944.72	12,944.72	0.0125

Discount rate = 4%

NPV = \$2517.182

Variance = \$20,359,090.9093

Standard deviation = \$4512.105

example, the cash flow in Period 3 depends upon the size of the cash flow in Periods 2 and 1. Given the cash flow in Period 1, the cash flow in Period 2 can vary within a range, but more of a known range than if we did not know the cash flow in Period 1. Similarly, the cash flows in Period 3 are allowed to vary within a range of values, the range being determined in part by the realized cash flow in the prior period. Most importantly, it should be noted that we assume we know what the distribution of each cash-flow figure is, and the probabilities associated with each flow, and, further, that once we make the investment decision we more or less have locked ourselves into the project. In the following section, we relax this latter assumption and allow sequential decisions to be introduced into the conditional-probability decision framework.

References

- Bonini, C.P. Spring 1975. Comment on formulating correlated cash-flow streams. *The Engineering Economist* 20: 269–314.
- Chen, S. N., and W. T. Moore, “Investment decision under uncertainty: Application of estimation risk in the Hillier approach,” *Journal of Financial and Quantitative Analysis* 17 (September 1982): 425–440.
- Hillier, F., “The derivation of probabilistic information for the evaluation of risky investments,” *Management Science* 9 (April 1963): 443–457.
- Wang, S.Y., and C.F. Lee. 2010. A fuzzy real option valuation approach to capital budgeting under uncertainty environment. *International Journal of Information Technology & Decision Making* 9 (5): 695–713.

Appendix M: Capital-Rationing Decision

In this section, we discuss a capital-budgeting problem that involves the allocation of scarce capital resources among competing economically desirable projects, not all of which can be carried out due to a capital (or other) constraint. This kind of problem is often called “capital rationing.” In this section, we show how linear programming can be used to make capital-rationing decisions.

Basic Concepts of Linear Programming

Linear programming is a mathematical technique used to find optimal solutions to problems of a firm involving the allocation of scarce resources among competing activities. Mathematically, the type of problem that linear programming can solve is one in which both the objective of the firm to be maximized (or minimized) and the constraints limiting the firm’s actions are linear functions of the decision variables involved. Thus, the first step in using linear programming as a tool for financial decisions is to model the problem facing the firm into a linear-programming form. To construct the linear-programming model, one must take the following steps.

First, identify the controllable decision variables involved in the firm’s problem. Second, define the objective or criterion to be maximized or minimized and represent it as a linear function of the controllable decision variables. In finance, the objective generally is to maximize the profit contribution or the market value of the firm or to minimize the cost of production.

Third, define the constraints and express them as linear equations or inequalities of the decision variables. This will usually involve (a) a determination of the capacities of the scarce resources involved in the constraints, and (b) a derivation of a linear relationship between these capacities and the decision variables.

Symbolically, then, if $X_1, X_2, \dots, X_n$ represent the quantities of output, the linear-programming model takes the general form:

$$\text{Maximize (or minimize) } Z = c_1X_1 + c_2X_2 + \cdots + c_nX_n, \quad (\text{A.84})$$

Subject to: $a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n \leq b_1,$
 $a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n \leq b_2,$
 $\vdots \qquad \qquad \qquad \vdots$
 $a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n \leq b_m,$
 $X_j \geq 0, (j = 1, 2, \dots, n).$

Here Z represents the objective to be maximized (or minimized), profit, or market value (or cost); $c_1, c_2, \dots$, and c_n and $a_{11}, a_{12}, \dots$, and a_{mn} are constant coefficients relating to profit contribution and input, respectively; $b_1, b_2, \dots$, and b_m are the firm's capacities of the constraining resources. The last constraint ensures that the decision variables to be determined are nonnegative.

Several points should be noted concerning the linear programming model. First, depending upon the problem, the constraints may also be stated with equal signs ($=$) or as greater-than-or-equal-to. Second, the solution values of the decision variables are divisible, that is, a solution would permit $x(j) = 1/2, 1/4$, etc. If such fractional values are not possible, the related technique of integer programming, yielding only whole numbers as solutions, can be applied. Third, the constant coefficients are assumed known and deterministic (fixed). If the coefficients have probabilistic distributions, one of the various methods of stochastic programming must be used. Examples will be given below of the application of linear programming to the areas of capital rationing and capital budgeting.

Capital Rationing

The XYZ Company produces products A, B, and C within the same product line, with sales totaling \$37 million last year. Top management has adopted the goal of maximizing shareholder wealth, which to them is represented by gain in shareholder price. Wickwire plans to finance all future projects with internal or external equity; funds available from the equity market depend on share price in the stock market for the period.

Three new projects were proposed to the Finance Committee, for which the following net after-tax annual funds flows are forecast:

Project	Year					
	0	1	2	3	4	5
X	−100	30	30	60	60	60
Y	−200	70	70	70	70	70
Z	−100	−240	−200	400	300	300

All three projects involve financing cost-saving equipment for well-established product lines; adoption of any one project does not preclude adoption of any other. The following NPV formulations have been prepared by using a discount rate of 12%.

Investment	NPV
X	65.585
Y	52.334
Z	171.871

In addition, the finance start has calculated the maximum internally generated funds that will be available for the current year and succeeding 2 years, not counting any cash generated by the projects currently under consideration.

Year 0	Year 1	Year 2
\$300	\$70	\$50

Assuming that the stock market is in a serious downturn, and thus no external financing is possible, the problem is which of the three projects should be selected, assuming that fractional projects are allowed.

The problem essentially involves the rationing of the capital available to the firm among the three competing projects such that share price will be maximized. Thus, assuming a risk-adjusted discount rate of 12%, the objective function becomes:

$$\text{Maximize } V = 65.585X + 52.334Y + 171.871Z + 0C + 0D + 0E$$

where V represents the total present value realized from the projects, and C , D , and E will represent idle funds in period 0, 1, and 2, respectively. The constraint for period 0 must ensure that the funds used to finance the projects do not exceed the funds available. Thus,

$$100X + 200Y + 100Z + C + 0D + 0E = 300.$$

In this constraint, C represents any idle funds unused in period 0 after projects are paid for. Similarly, for period 1 and 2,

$$\begin{aligned} -30X - 70Y + 240Z - C + D + 0E &= 70, \\ -30X - 70Y + 200Z + 0C - D + E &= 50. \end{aligned}$$

Here $-D$ and $-E$ are included in the second and third constraints, ensuring that idle funds unused from one period are carried over to the succeeding period. In addition, to prevent the program from repeatedly selecting only one project (the “best”) until funds are exhausted, three additional constraints are needed:

$$X \leq 1, \quad Y \leq 1, \quad Z \leq 1.$$

The solution to the model, if $V = \$208.424$, is:

$$X = 1.0, \quad Y = 0.6586, \quad Z = 0.6305.$$

To give an indication of the value of relaxing the funds constraint in any period (the most the firm would be willing to pay for additional financing), the shadow price of the funds constraints are given below:

Funds constraint	Shadow price
1st period	0.4517
2nd period	0.4517
3rd period	0.0914

It should be noted that the constraints related to $X \leq 1$, $Y \leq 1$, and $Z \leq 1$ are required in solving capital-rationing problems. If these constraints are removed, then we will obtain $X = 2.4074$, $Y = 0$, and $Z = 0.5926$. This issue has been discussed by Copeland and Weston (2003) and Weingartner (1963, 1977).

Thus, linear programming is a valuable mathematical tool with which to solve capital-budgeting problems when funds rationing is required. In addition, duality has been used by the banking industry to determine the cost of capital of funds.

References

- Copeland, T.E., and J.F. Weston. 2003. *Financial theory and corporate policy*. 4th ed. Reading: Addison-Wesley Publishing.
- Weingartner, H.M. 1963. *Mathematical programming and the analysis of capital-budgeting problems*. Englewood Cliffs: Prentice-Hall.
- . 1977. Capital rationing: N authors in search of a plot. *Journal of Finance* 32: 1403–1432.

Appendix N: Noncentral χ^2 and the Option Pricing Model

It is well known that $Y = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma_x} \right)^2$ is distributed as χ^2 with n degree of freedom.

This is a central χ^2 distribution. It can be shown that $Y' = \sum_{i=1}^n X_i^2$ is distributed as noncentral χ^2 with n degree of freedom and a noncentral parameter

$$\lambda = \frac{1}{2} \sum \mu_i^2.$$

If $\mu = 0$, the distribution of Y' reduces to the central χ^2 distribution.

The Black-Scholes Option Pricing Model assumed that the variance of stock rate of return (σ^2) is constant. If the variance of stock rate of return is a function of stock price per share, $\sigma^2 S^{\beta-2}$, then the Black-Scholes model can be generalized as¹

$$C = S \left[1 - \chi^2 \left(2n; 2 + \frac{2}{2-\beta}, 2m \right) \right] - Xe^{-r(T-t)} \left[\chi^2 \left(2m; \frac{2}{\beta-2}, 2n \right) \right] \quad (\text{A.85})$$

$(\beta < 2)$

$$C = S \left[1 - \chi^2 \left(2m; \frac{2}{2-\beta}, 2n \right) \right] - Xe^{-r(T-t)} \left[\chi^2 \left(2n; 2 + \frac{2}{\beta-2}, 2m \right) \right] \quad (\text{A.86})$$

$((\beta < 2))$

where T = time of expiration of option, t = current time, r = risk-free rate. $\chi^2(W, V, \lambda)$ is the cumulative noncentral chi-square distribution function with W , V , and λ being the upper limit of the integral, degree of freedom, and noncentrality, respectively. In addition, m , n , and K can be defined as:

$$m = KS^{2-\beta} e^{(2-\beta)\mu(T-t)}$$

¹The derivation of this formula can be found in Schroder (1989).

$$n = KS^{2-\beta}$$

$$K = \frac{2\mu}{\sigma^2(2-\beta)(e^{(2-\beta)\mu(T-t)} - 1)} \quad (\text{A.87})$$

Now, we discuss three possible special cases associated with Eqs. A.85 and A.86.

- (a) If $\beta = 2$, both m and n approach infinity. Then it can be shown that both Eqs. A.85 and A.86 reduce to the well-known Black-Scholes formula as defined in Appendix K.
- (b) If $\beta = 1$, it can be shown that Eqs. A.85 and A.86 reduce to

$$C = (S - Xe^{-r(T-t)})N(y_1) + (S + Xe^{-r(T-t)})N(y_2) + \nu[n(y_1) - n(y_2)]$$

where

$$\nu = \sigma \sqrt{\frac{1 - e^{-2r(T-t)}}{2r}}$$

$$y_1 = \frac{S - Xe^{-r(T-t)}}{\nu}$$

$$y_2 = \frac{-S - Xe^{-r(T-t)}}{\nu}$$

$N(y_1)$ and $N(y_2)$ = cumulative standardized normal distribution function in terms of y_1 and y_2 , respectively.

$n(y_1)$ and $n(y_2)$ = standardized normal density function in terms of y_1 , and y_2 , respectively.

- (c) If $\beta = 0$, it can be shown that Eqs. A.85 and A.86 reduce to

$$C = SN[q(4)] - Xe^{-r(T-t)}N[q(0)] \quad (\text{A.88})$$

where

$$q(w) = \frac{1 + h(h-1)\left(\frac{w+2y}{(w+y)^2}\right) - h(h-1)\left(2 - h(1-3h)\left(\frac{(w+2y)^2}{2(w+y)^4}\right) - \left(\frac{z}{(w+y)}\right)^h\right)}{\left\{2h^2\left(\frac{w+2y}{(w+y)^2}\right)\left(1 - (1-h)(1-3h)\left(\frac{w+2y}{(w+y)^2}\right)\right)\right\}^{1/2}}$$

$$h(w) = 1 - \frac{2}{3}(w+y)(w+3y)(w+2y)^{-2}$$

$$y = \frac{4rS}{\sigma^2(1 - e^{-r(T-t)})} \quad \text{and} \quad z = \frac{4rX}{\sigma^2(e^{-r(T-t)} - 1)}$$

The elasticity of variance ($\sigma^2 S^{\beta-2}$) with respect to stock price per share S is

$$\eta_s = \left[\frac{\partial(\sigma^2 S^{\beta-2})}{\partial S} \right] \left[\frac{S}{\sigma^2 S^{\beta-2}} \right] = \left[\frac{(\beta - 2)\sigma^2 S^{\beta-2}}{S} \right] \left[\frac{S}{\sigma^2 S^{\beta-2}} \right] = \beta - 2 \quad (\text{A.89})$$

This implies that the Option Pricing Model defined in Eqs. A.85 and A.86 is a constant elasticity of variance (CEV) type of OPM.

The CEV type of option pricing model can be reduced to the following special models.²

- (a) $\beta = 2$, Eqs. A.85 and A.86 reduce to the Black-Scholes model.
- (b) $\beta = 1$, Eqs. A.85 and A.86 reduce to the absolute model as defined in Eq. A.88.
- (c) $\beta = 0$, Eqs. A.85 and A.86 reduce to the square root model as defined in Eq. A.89.

Reference

- Schroder, M. 1989. Computing the constant elasticity of variance option pricing formula. *Journal of Finance* 44: 211–220.
- Beckers, S. (1980). The constant elasticity of variance model and its implications for option pricing. *The Journal of Finance* 35(3): 661–673.

²See Beckers (1980).

Appendix O: Impacts of Financing Decisions on Capital Budgeting Decisions

This appendix is intended to briefly review net-present value (NPV) method and explain how financing decision can affect investment decisions.

Chambers, Harris, and Pringle (1982) (CHP) examined four standard, discounted cash-flow models and considered the implications of using each as opposed to the other three. By way of simulation, they were able to deal with differences in financing projects as well as with possible differences in the risk underlying the operating cash flows generated by each project. The problem inherent in this and any project evaluation (not to say that there are no other difficulties) is in concentrating on the specification of the amount of debt used to finance the investment project. Project debt is therefore defined as the *additional debt capacity* afforded the firm as a result of accepting the project and can alternatively be described as the difference between the firm's optimal debt level *with* the project and *without* it.

The CHP analysis considered the following four methods; (i) the Equity Residual method; (ii) the “after-tax Weighted Average Cost-of-Capital” method; (iii) the “Arditti-Levy” weighted cost-of-capital method; and (iv) the Myers “Adjusted Present Value” method. For simplicity, Table A.6 contains the definitions of the variables used in the following discussion, after which we briefly discuss the formulation of each method, and then elaborate on the way in which each method incorporates interacting effects of financing and investment decisions.

Using the variables defined in Table A.6, we now explicitly define the equations for the four alternative methods.

Equity-Residual Method

The equity-residual method is formulated in Eq. (A.90) in a manner that emphasizes the goal of financial management, the pursuance of the interests of shareholders:

$$NPV(ER) = \sum_{t=1}^N \frac{[(R_t - C_t - dep_t - rD_t)(1 - \tau_c) + D_t] - (D_t - D_{t+1})}{(1 + k_e)^t} - [I - NP]. \quad (A.90)$$

Table A.6 Definitions of variables

R_t	= Pretax-operating cash revenues of the project during period t
C_t	= Pretax-operating cash expenses of the project during period t
dep_t	= Additional depreciation expense attributable to the project in period t
τ_c	= Applicable corporate tax rate
I	= Initial net cash investment outlay
D_t	= Project debt outstanding during period t
NP	= Net proceeds of issuing project debt at time zero
r_t	= Interest rate of debt in period t
k_e	= Cost of the equity financing of the project
k_w	= After-tax weighted-average cost of capital (i.e., debt cost is after-tax)
k_{AL}	= Weighted average cost of capital – debt cost considered before taxes
ρ	= Required rate-of-return applicable to unlevered cash-flow series, given the risk class of the project
r , k_e , and ρ are all assumed to be constant over time	

The formula presented above can be interpreted as stating that the benefit of the project to the shareholders is the present value of the cash flows not going to pay operating expenses or to service or repay debt obligations.

With these flows identified as those going to shareholders, it is appropriate, and rather easy, to discount these flows at the cost of equity. The only difficulty involved is identifying this cost of equity, a problem embodied in all capital-budgeting methods.

After-Tax, Weighted-Average, Cost-of-Capital Method

The after-tax, weighted-average, cost-of-capital method, depicted in Eq. (A.91), has two noticeable differences from the formulation of the equity-residual method:

$$NPV = \sum_{t=1}^N \frac{(R_t - C_t - dep_t)(1 - \tau_c) + dep_t}{(1 + k_w)^t} I. \quad (\text{A.91})$$

First, no flows associated with the debt financing appear in the numerator, or as relevant cash flows. Second, the cost of capital is adjusted downward, with k_w being a weighted average of debt and equity costs, the debt expense accounted for on an after-tax basis. In that way, debt financing is reckoned with an indirect manner. With the assumption that r and k_e are constant over time, k_w can be affected only by the debt-to-equity ratio, a problem most often avoided by assuming a fixed debt-to-equity ratio.

Arditti and Levy Method

The Arditti-Levy method is most similar to the after-tax weighted-average cost-of-capital method. This formulation can be written as:

$$NPV(AL) = \sum_{t=1}^N \frac{[(R_t - C_t - dep_t - rD_t)(1 - \tau_c) + dep_t] + rD_t}{(1 + k_{AL})^t} - I \quad (A.92)$$

It was seen as necessary to restate the after-tax, weighted-average, cost-of-capital formula in this manner because the tax payment to the government has an influence on the net cash flows, and for that reason the cash-flow figures would be misleading. To rectify this problem, the discount rate must now be adjusted since the interest tax shield is reflected in the cash flows and double counting would be involved. While the after-tax, weighted-average cost of capital recognized the cost of debt in the discount rate, it was akin to the equity-residual method in considering only returns to equity. The Arditti and Levy formulas imply that a weighted average discount rate including the lower cost of debt could only (or best) be used if all flows to all sources of financing were included; hence the term rD_t is found at the end of the first term. This can be rationalized if one considers the case of a firm where there is one owner. The total cash flow to the owner is the relevant figure, and the discount rate applicable is simply a weighted average of the two individual required rates of return.

Myers Adjusted-Present-Value Method

This method is closely related to the Arditti-Levy method except for the exclusion of the interest-expense flows to the bondholders. In treating the financing mix, Myers implicitly assumes that the tax shield, created by the interest payments and afforded to the equity holders, has the same risk for the equity holders as the coupon or interest payments have for the bondholders. Instead of aggregating all factors and attempting to arrive at a suitable weighted-average discount rate, Myers found it less difficult to leave operating- and financing-related flows separated, and to discount each at an appropriate rate. This formulation can be written as:

$$APV = \sum_{t=1}^N \frac{(R_t - C_t - dep_t)(1 - \tau_c) + dep_t}{(1 + \rho)^t} - I + \sum_{t=1}^N \frac{\tau_c r D_t}{(1 + r)^t} \quad (A.93)$$

This formulation appears to be closely related to the Modigliani and Miller with-tax firm-valuation model, and it should, given Myers' motivation for this work. We choose not to discuss it here, but the reader should be aware that Myers' emphasis was not solely on the tax advantage of debt, as the last term in the above equation tends to imply.

The four methods are obviously comparable but usually give different figures for the net present value. The equity-residual, after-tax, weighted-average cost of capital and the Arditti-Levy weighted-average, cost-of-capital formulations are comparable if the value of the debt outstanding remains a constant proportion of the remaining cash flows; this is often taken to mean a constant debt ratio. This will

not guarantee that the Myers APV method will yield the same results, although the Myers method is equivalent to the other three only if the project life can be described as one period, or as *infinite with constant perpetual flows* in all periods. By way of numerical examples, we now address the task of determining the factors that create the differences in the net-present-value figures generated by each method.

The first example involves the simplest case, where a \$1000 investment generates operating flows of \$300 per year for all 5 years of the project's life. Debt financing comprises \$600 of the project's total financing, and the principal is repaid over the five time periods in equal amounts. The discount and tax rates employed are included below in Table A.7 with the results for each of the four methods.

Because the Arditti-Levy method recognizes the acquisition of the debt capital and uses a lower discount rate, as does the after-tax WACC, these two methods give the highest net-present-value figures. The equity-residual method recognizes only the \$400 outflow at time zero, but the higher discount suppresses the net-present-value figure. The Myers APV method, though discounting financing-related flows at the cost of debt, also attains a low net-present-value figure because the majority of the flows are discounted at a higher unlevered cost-of-equity rate.

Basically, we can speak of three differences in these methods that create the large discrepancies in the bottom-line figures. The risk factor, reflected in discount rates that may vary over time, is a major element, as was evidenced in the example presented above. The pattern of debt and debt payments will also be an important factor, particularly so when the debt repayment is not of the annuity form assumed earlier. Finally, the recognition and valuation of the debt tax shields will play an important role in net-present-value determination, especially when the constant-debt-repayment assumption is dropped, and the interest expenses and associated tax shields grow larger.

Since the Myers APV method is recommended, it is worthwhile for us to show the procedure for obtaining the APV, as discussed in Table A.7. Let the information needed to establish APV be $R_t = \$4000$, $C_t = \$1000$, $dep_t = \$500$, $D_t = \$3600$, E_t (equity financing) = \$2800, $I = \$7000$, $k_e = 0.15$, $r = 0.09$, $\tau_c = 0.46$, $w = D/(D + E) = \$4200/\$7000 = 0.6$, and $N = 6$.

Table A.7 Application of four capital budgeting techniques

Inputs: 1. $k_e = 0.112$ 2. $r = 0.041$ 3. $\tau_c = 0.46$ 4. $\rho = 0.0802$ 5. $w = 0.6$		
Method	NPV results	Discount rates
1. Equity-residual	\$230.55	$k_e = 0.112$
2. After-tax WACC	270.32	$k_w = 0.058$
3. Arditti-Levy WACC	261.67	$k_{AL} = 0.069$
4. Myers APV	228.05	$r = 0.041$ and $\rho = 0.0802$

From this information, we can assemble the information needed to estimate APV as follows:

$$\begin{aligned}\text{Operating flows} &= (R_t - C_t - dep_t)(1 - \tau_c) + dep_t \\ &= (\$4000 - \$1000 - \$500)(1 - 0.46) + \$500 \\ &= \$1850.\end{aligned}$$

$$\begin{aligned}\text{Financial flows} &= \tau_c r D_t \\ &= (0.46)(0.09)(\$3600) \\ &= \$149.04\end{aligned}$$

$$\begin{aligned}k_w &= wr(1 - \tau_c) + (1 - w)k_e \\ &= (0.6)(0.09)(1 - 0.46) + (1 - 0.6)(0.15) \\ &= (0.029) + (0.06) \\ &= 0.089.\end{aligned}$$

$$\begin{aligned}\rho &= \frac{k_w}{(1 - \tau_c w)} \\ &= \frac{0.089}{(1 - (0.46)(0.6))} \\ &= 0.1229\end{aligned}$$

From this information, we can calculate APV as:

$$\begin{aligned}APV &= \$1850 \left(\sum_{t=1}^6 1/(1 + 0.1229)^t \right) + \$149.04 \left(\sum_{t=1}^6 1/(1 + 0.09)^t \right) - \$7000 \\ &= (\$1850)(4.0784) + (\$149.04)(4.4859) - \$7000 \\ &= \$1213.62.\end{aligned}$$

By using appropriate budget constraints, Lambrecht and Myers (2012) have theoretically shown that investment dividend and financing policy are interrelated. To test this theory developed by Lambert and Myers, Lee et al. (2016) investigated the interrelationship among investment, financing, and dividend decisions using 2SLS, 3SLS, and GMM methods based on the US listed firm annual data between 1965 and 2012. Their results were consistent with Lambrecht and Myers' (2012) theory that dividend and investment decisions are jointly determined. In addition, these three corporate decisions are codetermined and the interaction among them should be taken into account in a simultaneous equation framework.

References

- Beckers, S. 1980. The constant elasticity of variance model and its implications for option pricing. *Journal of Finance* 35: 661–673.
- Chambers, D. R., R. S. Harris, and J. J. Pringle, “Treatment of financing mix in analyzing investment opportunities,” *Financial Management* 11 (Summer 1982): 24–41.
- Lambrecht, B.M., and S.C. Myers. 2012. A Lintner model of payout and managerial rents. *Journal of Finance* 67: 1761–1810.
- Lee, Cheng-Few, Woan-Lih Liang, Fu-Lai Lin, and Yating Yang. 2016. Application of simultaneous equation in finance research: Methods and empirical results. *Review of Quantitative Finance and Accounting* 47 (4): 943–971.

Index

A

- AAUC, 2198, 2199
- ABC bond corporation, 1039, 1040
- Abnormal returns, 890, 1181–1183
- Abnormal sales growth, 2502, 2504
 - and abnormal returns, 2507–2508
 - alternative econometric estimate methods, 2530
 - background and hypotheses development, 2504
 - correlation results, 2514–2515
 - empirical test methods, 2508
 - Fama-MacBeth regressions, 2520–2524
 - firm characteristics based on rank of, 2530, 2531
 - firm-level abnormal return, 2530
 - firm-specific effect, 2529
 - future returns, 2508
 - industry and year effect, 2529
 - level of operating risk, 2533
 - ordinary least square regressions, 2518–2520
 - ranked portfolio returns, 2524–2528
 - sample selection and data, 2510–2512
 - sample summary statistics, 2512–2513
 - unexpected sales growth and volatility of sales, 2530
- “Abnormal” total accruals (ΔTAt), 2505
- Abnormal vibrations, 1075
- Absolute Priority Rule (APR), 1302
- Absolute risk aversion (ARA), 1943, 1955
 - for favorable outcomes, 1953
- Accounting, 879, 887–888
 - conservatism, 1219
 - scandals, 1018, 1024
 - standards, 2541, 2549, 2552
- Accounting Standard Codification (ASC), 2026, 2493
- Accounts payable, 658, 662–664
- Accounts receivable, 652, 654–658
- Accrual(s), 2074
 - accounting system, 2504
- Accruals-based earnings management, 2064–2066, 2069, 2071–2074, 2081, 2085–2087
- Accruals quality, 2502, 2504, 2534
 - background and hypotheses development, 2504
 - correlation results, 2514–2515
 - empirical test methods, 2508
 - Fama-MacBeth regressions, 2520–2524
 - innate factors affecting, 2506
 - OLS regression, 2518–2520
 - sample selection and data, 2510–2512
 - summary statistics, 2513
- Acquisition, 878, 2460
- Activism, 968
- Actual exchange-rate, 1147
- Actual wage growth, 1147
- Actuarial assumptions, 2044
- Adaptive-expectation model, 1605
- Adjustable rate mortgage (ARM), 2264, 2265, 2271
- Adjusted Jensen alpha, 765–768
- Adjusted present value (APV), 2488, 2490, 2491
 - method, 1642, 1655
- Adjusted R-squares, 2440
- Administration order, 776
- Administrative receivership, 777
- Administrative vs. accounting controls, 1206
- Admissible performance, 670
- Adverse selection, 1009
- Affine jump diffusion (AJD), 1081–1082
- Agencies credit efficiencies, 2483
- Agency costs, 730, 883
- Agency problem, 1189
- Agency theory, 879, 881–883

- Agent-based models of financial markets
 - advantages, 1029
 - analytical tractability, 1028
 - autonomous agents, 1030
 - classifiers, 1031
 - complex adaptive systems, 1027
 - computer simulations, 1028
 - decision-making, 1028
 - decision making rules, 1031
 - design questions, 1033
 - endogenous transitions, 1027
 - heterogeneous agent rational expectations models, 1028
 - heterogeneous agents, 1030
 - inductive reasoning/induction, 1030, 1031
 - myopic/intertemporal preferences, 1030
 - one risk-free asset, 1032
 - research objectives, 1029
 - stock market, 1027
 - trading mechanism, 1032, 1033
 - traditional equation-based models, 1028
 - zero intelligence agents, 1030
- Aggregate shock, 1358
- Aggregator function, 557
- Aggressive accounting
 - competitive environment and incentives for, 1018–1020
 - hedging transactions, effectiveness of, 1020–1021
 - unconsolidated entities, control and risks relating to, 1021–1022
- Aggressive approach, 665
- All-Cash Transactions, 2479
- Alliance Mutual Fund, 1279
- Allias paradox, 850, 851
- Alpha, 640, 642–644, 647, 648
- Amazon.com, 756, 903
- American Express (AXP), 1991
- American Stock Exchange, 1095
- Amortization, 2264
- Analysts, 968
- Anam, K., 1398
- Annual-earnings announcement, 941
- Annual percentage rate (APR), 2264
- Anti-greenmail, 1211
- Antitakeover, 732, 734–737
- Antitakeover defenses, 1211
 - Charter and Bylaws Provisions, 2476, 2477
 - disadvantages, 2478
 - Fortune 500 companies, 2474
 - global financial crisis, 2474
 - golden parachutes, 2477, 2478
 - measures, 2474
 - Pac-Man, 2477
 - poison pill, 2475, 2476
 - re-actions, 2477
 - unknown threats, 2474
 - white knight and white squire, 2478
- Antitrust, 2459, 2460, 2466, 2472, 2474, 2482, 2496
- APT risk model, 2287–2292
- Arbitrage and market frictions
 - exact replication and prices under no friction, 954
 - no short sales, 955–956
 - payoffs, 953
 - simple binomial model, 956–958
 - trading, 953
- Arbitrage-free condition, 1488, 1490, 1501, 1504
- Arbitrage-free hedging, 828
- Arbitrage-free models, 1914
 - arbitrage-free hedging, 828
 - binomial lattice, 829
 - Black–Derman–Toy model, 832
 - Black–Scholes model, 827, 828
 - expected return, 831
 - Ho–Lee model, 827, 829
 - Hull–White model, 832, 833
 - local expectation hypothesis, 831
 - normal distributions, 831
 - one-period forward rate, 830
 - recombining condition, 828
 - standard economic theory, 827
 - Vasicek model, 829
 - zero-coupon bond, 830
- Arbitrage models, 867–868
- Arbitrage pricing model, 614, 622–624
- Arbitrage pricing theory (APT), 556, 561, 622–624, 671, 749, 868
- ARCH/GARCH jump diffusion model, 1087–1089
- ARCH GARCH Model, 1705
- ARMA-EVTGARCH, 2194–2197, 2199–2202
- ARMA-GARCH, 2192–2197, 2199–2202
- ARMA model, 1990
- Arm’s length financing, 792
- AROC, 2198
- Arrow–Pratt coefficient, 562
- Artificial intelligence (AI), 1031, 2095–2097
- ASC 842, 2026–2029, 2031, 2033–2035, 2037, 2038
- Asian Basket Currency (ABC), 1038, 1039
- Asian Bond Market Initiative (ABMI), 1038, 1039

- Asian bond markets
 - Asian economies, 1052, 1053
 - borrowing costs, 1052
 - concept, 1037
 - development, 1037
 - domestic corporate bonds, 1053
 - emerging markets, 1052
 - forms, 1038
 - fund, 1037
 - government revenues spent, 1052
 - local, 1054
- Asian corporations, 1037
- Asian crisis in 1997, 1181
- Asian Development Bank (ADB), 1041
- Asian economies, 1043, 1052
- Asian finances
 - Asian financial hubs, 1049
 - Asian regulators, 1048, 1049
 - capital flows, 1047
 - capital importer/exporter, 1046
 - financial market stability, 1048, 1049
 - geopolitics, 1049
 - investor motivations, 1047, 1048
- Asian financial crisis, 1043
- Asian financial hubs, 1049
- Asian options, 1887
- Asian Securitization and Infrastructure Assurance Ltd (ASIA Ltd), 1041
- Asset allocation, 866
 - strategy, 1986
- Asset-based approach, 2493
- Asset-based lending, 2472
- Asset Management Company, 1282
- Asset market approach, 947
- Asset price bubble, 810
- Asset pricing, 954, 959, 1345, 1357–1359
- Asset pricing model(s), 560, 672, 716, 744, 1472–1474
 - arbitrage opportunities, 613
 - arbitrage pricing model, 622–624
 - CAPM, 617
 - consumption-based, 617–619
 - consumption expenditures, 615
 - Euler equations, 559–560
 - expected excess returns, 616
 - factor models, 622–624
 - financial assets, 613
 - financial economics, 624
 - financial market equilibrium, 614
 - human imperfections role, 614
 - IMRS, 615
 - log-linear budget constraint, 558–559
 - mean–variance efficiency, 621
 - multi-beta, 619–621
 - “no arbitrage principle”, 613
 - nonexpected utility, 557–558
 - SDF, 614–616
 - security returns to aggregate quantities, 616
 - systematic risk, 616
- Asset swaps, 574
- Asymmetric dynamic conditional correlation estimates, 2574
- Asymmetric timeliness
 - accounting variable collection, 1844
 - accrual and cash flow components, 1830, 1840, 1841
 - accrual components, 1856
 - aggregated earnings, 1830
 - application, 1855, 1857
 - asymmetric verification requirement, 1832
 - clean surplus earnings, 1852, 1861, 1863
 - conduct comparative analysis, 1843
 - conservatism, 1863
 - contingent assets, 1833, 1857, 1858
 - contingent liabilities, 1833
 - DATASTREAM, 1843
 - decomposition, line items, 1839, 1840
 - decomposition, 1831, 1863
 - descriptive statistics and correlations, 1847
 - dimensions, 1830
 - dirty surplus flows, 1858–1860
 - earnings, 1863
 - earnings components, 1837–1839, 1841, 1842
 - earnings conservatism, 1830
 - economic consequences, 1830
 - efficient technology, 1832
 - financial reporting, 1830
 - impairment test, 1833, 1857, 1858
 - information asymmetry, 1831
 - informativeness, 1830
 - innovative feature, 1830
 - LCM rule, 1831, 1857
 - liabilities, 1857, 1858
 - limitation, 1832, 1834
 - lower of cost or market value, 1832
 - operating cash flow and accrual components, 1851, 1854, 1855
 - pattern, 1830
 - reverse regression, 1830
 - role of conservatism, 1830
 - sample collection and variables, 1846
 - screening processes, 1846
 - sensitivity, 1832
 - shareholders, 1831
 - sign of market returns, 1837

- Asymmetric timeliness (*cont.*)
 - statistics, 1846, 1851
 - surplus earnings, 1831
 - UK financial statements, 1863
- Asymmetric upward and downward jumps, 1089
- AT&T spin off NCR, 2471
- AUC, 2189, 2198–2200
- Audit committee, 1023, 1196, 1198, 1200, 1210, 1217
 - disclosures, 1201
 - establish, 1199
 - financial expertise, 1201
 - functions, 1199
 - independence, 1200
- Auditing research, 2106–2108
- Auditors, independence of, 1210
- Audit quality, 2108
- Augmented CAPM, 766–768
- Augmented Dickey-Fuller (ADF), 1716, 1724
- Augmented Dickey Fuller (ADF) test statistics, 1422, 2563
- Autocorrelation, 940, 941
- Auto-covariance, 1111
- Autoregressive Conditional Heteroscedastic (ARCH), 586
- Autoregressive jump process model, 1083–1084
- Auto-regressive models
 - ARMA-EVTGARCH, 2194
 - ARMA-GARCH, 2192–2194
 - CARL-vol model, 2195–2196
 - LPA, 2196
 - model selection, calibration and exceedance probability, 2194–2195
- Available stable funding (ASF), 2254
- Average sales price, 926
- Aversion parameters, 1587
- Axioma Robust Risk Model, 2294
- Axioma Statistical Risk Model, 2294
- Axioma World-Wide Equity Risk Factor Model (AX-WW2.1), 2294
- B**
- Babylonian Talmud, 866, 867
- Back-end statutory merger, 2469
- Backtest performance
 - historical performance, 2203–2205
 - risk profile, 2200–2202
 - switching strategy, 2205–2207
- Backward-looking approach, 1984
- Balance sheet enhance return strategies, 1913
- Balance sheet hedging, 696
- Balance-sheet identity, 1126
- Bank-base financial system, 798
- Bank branches, 1780–1782, 1785, 1796
- Bank-centered financial system, 796
- Bank credit, 1125, 1128
- Bank debt, 795, 797, 813
- Bank debt ratio, 793, 798, 802, 804, 806–808, 812
- Bank financing, 794, 797
- Bank-firm relationship, 796
- Bank for International Settlements (BIS), 1239
- Banking crisis, 513, 525
- Banking systems, 1987
- Bank of America
 - bankruptcy risks, 2003
 - capital ratios, 1996, 1997, 2005, 2011
 - risk-free assets, 1997
- Bank of International Settlements (BIS), 1038, 1984
- Bank regulations, 537
- Bank runs, 512–515
- Bankruptcy, 774, 775, 778, 779, 1299, 1302, 1310, 1312, 1315, 1766–1768, 1771, 1774–1777
 - avoidance hypothesis, 879, 886–887
 - estimation of, 2103
 - risk, 1989, 1997
- Bankruptcy prediction, 1767
 - accuracy rate, 1768
- Bank(s), 682, 685, 1780–1782, 1796
 - lending, 2255, 2256
 - profit function, 1128
 - stock returns, 1180
 - supervision, 515
- Barra risk model, 2279–2287
- Barra US-E3 factor index definitions
 - currency sensitivity, 2306
 - dividend yield, 2306
 - earnings variability, 2304, 2305
 - earnings yield, 2304
 - growth, 2302, 2304
 - leverage, 2305
 - momentum, 2300
 - non-estimation universe indicator, 2306
 - size, 2301
 - size nonlinearity, 2301
 - trading activity, 2301, 2302
 - value, 2304
 - volatility, 2299, 2300
- Barrier options, 1642, 1651, 1652
- Basel Capital Accord*, 2127

- Basel Committee on Banking Supervision (BCBS), 1239, 2252
- Basel I, 2252, 2253
- Basel II, 2253
- Basel III, 2258, 2259
 - available stable funding, 2254
 - capital regulation, 2255, 2256
 - liquidity coverage ratio, 2254
 - liquidity regulation, 2257
 - net stable funding ratio, 2254
 - required stable funding, 2254
- Basel standard, 1987
- Basket default swaps, 581–583
- Basu model, 1851
- Bayesian approach, 2385–2386
- Bayes' Theorem, 1310
- BC performance, 673
- Behavioral economics research with weighted probabilities, 1948
- BEKK model, 2558
- Belt and Road Initiative, 1050
- Benchmark comparison, 984
- Benchmark portfolio, 640, 644
- Benefield, J.D., 1398
- Bernoulli approximation, 1080
- Bernoulli jump process, 1077
- Best bid and offer (BBO), 1098, 1100
- Best clientele (BC), 673
- Beta pricing models, 615, 619, 621, 623, 624
- Bid-ask spread, 720, 721, 756, 1002, 1008, 1009
- Big data, and MapReduce systems, 1879–1880
- Binomial stock price model, 958
- Biotechnology business, 904
- Black, F., 1398
- Black and Cox (BC) model, 1299–1302
- Black–Derman–Toy model, 831, 832
- Black model, 748
- BlackRock Strategic Income Fund, 1269
- Black–Scholes equation, 902
- Black–Scholes formula, 1076
- Black–Scholes–Merton (BSM) model, 1294–1297, 1733
- Black–Scholes model, 827–830, 1419, 1587
- Black–Scholes pricing model, 1514
- Bloomberg's benchmark curves, 690
- Board directors, 969
- Board independence, 1192, 1196
 - CEO duality, 1195
 - negative aspects, 1193
 - ownership structure, 1194
 - positive aspects, 1193
- Board members, numbers, 1204
- Board's effectiveness, 1203
- Board size, 1197
- Bond duration, 683
 - vs. price volatility, 684
- Bond interest rate, 1127
- Bond markets, 1036
- Bond-portfolio management strategy, 682
- Bond price, 1729, 1730, 1739, 1741–1743, 1759, 1761
 - volatility, 685
- Bond pricing, 1306
- Bond yields, 1070
- Book building, 998, 1000
- Book income, 2540, 2542, 2543, 2546, 2548, 2550, 2552, 2553
- Book-tax conformity, 2540, 2542, 2548–2552
- Book-tax difference
 - accounting and finance studies, 2540
 - accounting income and taxable income, 2541
 - aggressive accounting/tax reporting practices, 2540
 - aggressive financial reporting, 2541
 - aggressive reporting practices, 2546
 - aggressive tax reporting, 2541, 2543, 2544
 - binary categorization, 2549
 - credit ratings, 2547
 - data availability, 2545
 - debate arises, 2548
 - definition, 2542
 - divergences, 2540
 - earnings, 2541
 - earnings management, 2544, 2550, 2552
 - earnings manipulation, 2540
 - earnings quality, 2546, 2547
 - economic activity, 2551
 - economic performance, 2542, 2552
 - effect of tax avoidance, 2546
 - effects, 2550
 - environment, 2550
 - equity, 2546
 - equity investors, 2551
 - ERC, 2551
 - financial accounting earnings, 2548
 - financial and tax reporting, 2545, 2546
 - financial reporting, 2541, 2551
 - financial statements, 2542, 2544, 2545, 2547, 2549
 - financial/tax reporting, 2540
 - firms, 2544
 - implementation, 2548
 - interpretation, 2545

- Book-tax difference (*cont.*)
 jurisdictions, 2549
 managers, 2541
 managers' aggressive reporting, 2541
 market participants, 2551
 measurement, 2541, 2544
 policymakers, 2548
 RMSE, 2549
 taxable income, 2541, 2542, 2545
 tax aggressiveness, 2544
 tax avoidance, 2540, 2544, 2547, 2550, 2552
 tax jurisdiction, 2545
 tax law, 2549
 tax liability, 2540, 2546, 2548
 tax returns, 2542, 2545, 2549, 2552
 timing difference, 2543
 transactions, 2542
 United States, 2540
- Book-to-market (BM), 676, 1111, 1117
- Bootstrap method, 1886
- Borrowers' credibility, 888
- Borrowers' debt structure, 793
- Borrowing constraints, 953
- Boston Stock Exchange, 1097
- Boundedly rational agents, 1027
- Break-up fees, 2479
- Breeden's model, 617
- Brennan and Schwartz model, 827, 833
- Brennan and Schwartz Two-Factor Model, 826
- Brownian motion, 1075, 1082–1085, 1576, 1885
- BTC, 2179, 2180, 2183, 2186, 2192, 2195, 2198, 2204, 2205, 2209
- Bulk purchase, 1318
- Bureau of Labor Statistics (BLS), 603
- Business corporation, 730
- Business judgment, 736
- Busy director, 1204
- C**
- Calculus model, 1565, 1566
- Calibration, 1905, 1911, 1915
- California Public Employees' Retirement System (CalPERS), 1021, 1209
- Call auction trading, 1014
- Call market trading, 1014
- Call option, 1882
- Campbell–Shiller model, 757
- Capital Asset Pricing Model (CAPM), 2485
- Capital asset pricing model (CAPM), 556, 561, 566, 567, 617, 620–622, 631–632, 634–635, 640, 671, 714, 747–748, 762–766, 859, 867, 871–874, 901, 902, 908–910, 928, 930, 933, 986, 2278, 2279, 2377, 2394–2402, 2489, 2490
- Bayesian approach, 2386
- grouping method, 2398
- instrumental variable method, 2399
- maximum likelihood method, 2401
- Capital asset pricing theory, 2416, 2428
- Capital budgeting, 1398, 1407, 2656, 2658
- DCF-based method, 2657
- decision, 2344
- fuzzy approaches, 2656, 2674
- and investment decision making, 2657
- methodologies, 2124
- traditional approaches, 2656
- under uncertainty, 2656, 2658
- Capital expenditure, 806
- Capital gain/loss, 687
- Capital investment decisions, 2038
- Capital investments, 2039
- Capital leases, 1673
- Capital market equilibrium, 1150, 1154
- Capital market line (CML), 1399, 2278
- Capital regulation, 2255
- Capital structure, 2402–2403
- arbitrage, 1257
- decision, 2042, 2048
- determinants of, 2371
- CARL-vol, 2195–2197, 2199–2202
- CAR model, 1182
- Cash, 656
- collections, 1668
- index volatility, 1412
- management, 651, 652, 654–655
- pay, 2142, 2144, 2146, 2149
- payments, 1668
- Cash-based accounting system, 2504
- Cash conversion cycle (CCC)
- aggressive liquidity policy, 2598
- cash flow management, 2598
- corporate governance, 2621
- corporate performance, 2602, 2606, 2608, 2612–2614
- COVID-19 pandemic, 2598
- data, 2600
- divergence, 2599
- economic development status, 2619, 2620
- endogeneity and controlling, 2622
- finance theory, 2598

- financial constraints, 2621, 2622
- financial crises, 2620
- financial market friction, 2599
- firm performance, 2599, 2613, 2615, 2622
- GMM, 2618
- legal protection, 2599
- liquidity management, 2599
- long-term financial decisions, 2622
- loss of business, 2599
- macroeconomic environments, 2599, 2619
- methodology, 2600–2602
- production process, 2599
- profitability and value, 2612
- regression, country, 2602, 2609, 2610
- regression, industry, 2610, 2611
- sample description, 2602, 2606, 2608
- 3SLS, 2615–2618
- three-stage least squares, 2617
- Cash flow (CF), 651–656, 659, 661, 662, 1116, 1586–1588, 2657, 2659, 2663, 2664, 2667, 2671, 2674
 - cycle, 662–664
 - distribution, 1585
 - payment rules, 1393–1394
- Cash flow statements
 - analysis, 1658
 - deferred employee compensation, 1678
 - direct method, 1662
 - direct vs. indirect method controversy, 1674–1676
 - dividends, 1679
 - financing activities, 1661
 - financing/securitization of payables, 1677, 1678
 - flexibility, 1662
 - free cash flows, 1682
 - income taxes, 1679, 1680
 - indirect method, 1662
 - investing activities, 1661
 - investing vs. financing activities, 1673
 - manipulation of payables, 1677
 - marketable securities, 1679
 - net income, 1658
 - non-cash transactions, 1681
 - operating vs. financing activities, 1672, 1673
 - operating vs. investing activities, 1671, 1672
 - sale and purchase of assets, 1680
 - SFAS 95, 1659
 - stock buybacks, 1678, 1679
- Catastrophic market risk, 1986
- Cauchy-Schwartz inequality, 1476
- Central limit theorem (CLT), 1985
- CEO compensation, 2140, 2152, 2154, 2156, 2158
 - comprehensive analysis, 2141
 - corporate finance, 2141
- CEO dismissal risk, 2141
- CEO duality, 1195
 - institutional investors, 1196
 - negative aspects, 1195
 - positive aspects, 1195
- CEO gender, 1686, 1689–1691, 1694, 1696–1698
- CEO labor market condition, 2149, 2165, 2167
- CEO marital status, 1689
- CEO risk preference, 2141, 2142, 2144, 2158, 2170
 - groups, 2149
- CEO's hiring contracts, 2141
- CEO's risk aversion, 2141
- CEO's risk preference, 2158
- Certainty effect, 851
- Certainty equivalent, 1399, 1402–1404, 1407
- Charter and Bylaws Provisions, 2476, 2477
- Chief Executive Officer (CEO), 1195
- China
 - B-share market, 714
 - equity markets, 713
 - foreign currency deposits, 714
 - foreign stock ownership, 714
 - ownership restriction, 714
 - SHSE, 713
 - stock shares, 713
 - SZSE, 713
- Chinese A and B shares
 - 1987–1998, 716
 - asset-pricing models, 716
 - asymmetric information, 715, 716
 - bid-ask spread, 717
 - CAPM, 714
 - cash flows, 715
 - companies (1990–1998), 716
 - domestic shares, 714
 - early- and mid-1990s, 714
 - foreign investors, 714
 - foreign shares, 714
 - London Stock Exchange, 715
 - Renminbi Special Shares, 714
 - return volatility, 715, 716
 - risk and return, 715
 - SHSE, 714
 - SZSE, 714
 - time-series features, stock returns and volatility, 714

- Chinese A and B shares (*cont.*)
 trading risk, 716
 trading summary, 11/2000–11/2001, 715
 transaction costs, 715, 717
 volatility, 716
- Chinese stock markets, 714, 1535
- Choi model, 1286
- Chu, P.K., 1398
- Chua, C.T., 1398
- CIR model, 826
- Cisco, 903
- Cisco System (CSCO), 1991
- Citigroup
 capital ratios, 2003, 2007, 2013
 estimation, 1996
- Classical estimation method
 constrained classical method, 2376
 multiple regression analysis, 2374–2376
 simple regression analysis, 2373–2374
- Class priority, 1095
- Clean surplus earnings, 1839
- Clientele specific fund performance, 671
- Closed-form solution, 898, 907, 908, 920, 932
- Cloud computing, 1871, 1872, 1874
- Cloud-enabled hybrid systems, 1880–1881
- Cluster computing, 1871
- Coarsened Exact Matching (CEM), 1784, 1786, 1788, 1789, 1792–1796
- Cobb-Douglas production function, 1141
- Coca-Cola Bottling, 2043
- Coefficient of relative risk aversion (CRRA), 1352
- Coefficient of variation, 1587
- Cointegration and error correction method (CECM)
 cointegration analysis, 1714
 conventional regression model, 1716
 error correction estimates, 1717
 error correction model, 1714, 1715
 heteroscedastic errors, 1716
 summary statistics, 1715
 techniques, 1714
 unit roots and residual cointegration tests, 1717
- Collateral, 703, 704, 707
- Collateralizability, 958
- Collateralized bond obligation (CBO), 1041
- Collateralized debt obligations (CDO), 1997
- Collateralized mortgage obligations, 1170
- Collateral loan obligations (CLOs), 2473
- Collective bargaining, 1138, 1139, 1147
- Collin-Dufresne and Goldstein (CDG) model, 1304, 1306
- Collins, B., 1398
- Committee on Foreign Investment in the United States (CFIUS), 2482
- Commodity Trading Advisors (CTAs), 1258
- Common stock, 1295, 1296
- Common Trend, 1781, 1782, 1785, 1786, 1788, 1790–1797
- Co-movement of stock market returns, 2560
- Company purchases, 888
- Company's attributes, 731
- Company voluntary arrangements, 776
- Compensation committees, 1202
 financial expertise, 1202
 functions, 1203
 procedures, 1203
 responsibilities, 1202
- Competency hypothesis, 1690
- Competitive dealer market, 1013
- Competitive environment, 1009
- Competitive IPO pricing, 2459
- Complex adaptive systems, 1027
- Composite earnings forecasting variable
 components, 2284
- Compound option approach, 1306
- Compulsory liquidation, 778
- Compute intensive IT systems, 1874–1877
- Computer assisted trading system (CATS), 1015, 1095
- Computerized Automated Execution System (CAES), 1098
- Conditional alpha, 640–642, 647
- Conditional asset pricing, 630
 conditional CAPM, 631–632
 definition, 629
 multi-beta conditional asset pricing models, 635–636
 return predictability, evidence for, 632–634
 tests of conditional CAPMs, 634–635
- Conditional beta, 641
- Conditional characteristic function (CCF), 1082
- Conditional covariance, 631
- Conditional distribution, 2637
- Conditional expectations, 630, 632
- Conditional jump dynamics model, 1085–1087
- Conditional market timing, 643–644
- Conditional maximum likelihood estimator (CMLE), 785
- Conditional performance, 669, 670

- Conditional performance evaluation (CPE),
639–641
conditional market timing, 643–644
conditional weight-based performance
measures, 644–646
empirical evidence, 646–648
implementations of, 641
- Conditional prepayment rate (CPR), 2271, 2274
- Conditional probability analysis (CPA),
781–783
- Conditional variance, 589, 590, 2559
- Conditional weight-based performance
measures, 644–646
- Confidence, 1582
- Conglomerate mergers, 2462
- Conservatism bias, 1114
- Conservative approach, 665
- Consolidated market, 1014, 1097, 1101,
1102, 1107
- Consolidated Quote System (CQS), 1098, 1099
- Constant absolute risk aversion (CARA),
2634, 2635
simple portfolio problem, 2642–2645
- Constant conditional correlation (CCC) model,
2569, 2573
- Constant Elasticity of Variance (CEV) diffusion
process, 1166
- Constant marginal personal tax rates, 2049
- Constant Proportion Portfolio Insurance
(CPPI), 1441–1443, 1445, 1447,
1449, 1455–1463, 2198
modification, 1449–1451
structured products, 1452
- Constant relative risk aversion (CRRA), 1585
- Consumer price index (CPI), 602, 610
- Consumer price index for all urban consumers
(CPI-U), 603, 604, 607
- Consumption-based asset pricing models
(C-CAPM), 617–619, 1350
- Consumption expenditure data, 618
- Consumption opportunity set
asset pricing model, 557–560
empirical evidence, 561
international asset pricing model,
567–570
Kreps–Porteus preferences, 562
optimal consumption and portfolio
allocation, 563–567
- Contingent immunization, 688, 689
- Continuous market, 1014
- Continuous-time option pricing methodology,
1159, 1161
- Continuous-time process, 1651
- Continuous trading, 992, 998–1000
- Continuous wavelet transform (CWT), 2580
- Contract devices, 970
- Contractual efficiency, 731, 732, 734, 738
- Controlled firms, 1194
- Control premium, 2492
- Control variables, 810
- Conventional regression model, 1716, 1723
- Convertible bonds, 583
- Convertible-security valuation, 1570
- Convexity of assets, 698
- Convexity of liabilities, 698
- Cornell model, 757
- Corporate and municipal zero-coupon, 543
- Corporate arbitrage
advantages of, 1330, 1331
capital structure proposition, 1335
equilibrium regimes of dominant
arbitrageurs, 1329
investor arbitrageurs, 1321, 1324
large investors versus, 1334
limits to investor arbitrage, 1319, 1320
shareholders of corporations, 1324–1329
- Corporate board(s)
directors, 1192
functions
audit committee, 1198 (*see also* Audit
committee)
committee structure, 1198
- Corporate bonds, 633, 695, 1295
- Corporate chartering system, 729, 730
- Corporate cultures, 881
- Corporate decisions, 2334
- Corporate disclosures, 1220
- Corporate diversification, 885
- Corporate dividend policy, 2410
- Corporate failure, 1192
causes of bankruptcy, 774–775
CPA, 781–783
discriminant analysis, 778
dynamic modeling, 787
financial ratio analysis, 778
LP, PM, and LM, 783–785
methods of bankruptcy, 775–778
optimal cut-off point, 785–787
static analyses, 787
- Corporate finance, 2141
- Corporate frauds, 1190
- Corporate governance, 1188, 1189, 2072
business judgment rule, 2480
fiduciary duties, 2480

- Corporate governance (*cont.*)
 - importance, 1190
 - merger activity, 2479
 - need, 1190
 - role, 1022–1024, 2480
 - seminal court cases, 2480, 2481
 - SOX, 1190, 1191
 - structure, 1191
- Corporate Governance Nominating Committee, 1201
 - annual report, 1202
 - composition, 1201
 - meetings/attendance, 1202
 - responsibilities, 1202
- Corporate innovation, 792, 815
- Corporate laws, 730–733, 735, 737–739
- Corporate loans, 703–705
- Corporate pension funding policy, 2048
- Corporate pension policy, 2048, 2054
- Corporate social responsibility (CSR), 2324, 2331
- Corporate tax rate, 2049
- Corporate tax shield, 2054–2056
- Corporate working capital
 - cash management, 654–655
 - money, 653–654
- Corporation, 1189
- Corrado and Miller (CM) formulas, 1521
- Corrected dynamic conditional correlation (cDCC), 2566, 2569
- Correlation coefficient, 1694
- Correlation matrix, 823
- Corresponding integral, 2053
- Cost mitigation
 - board oversight, 969
 - contract devices, 970
 - financial institutions, 969
 - government laws and regulations, 968
 - ownership structure, 968
 - securities traders, analysts, and press, 968
 - signaling mechanism, 970
- Cost of capital, 963, 970, 2370, 2387–2394
 - Bayesian approach, 2386
- Cost of equity capital, 1004
- Coupon rate, 683
- Covariances, 744, 745, 748, 750
- Covid-19 crisis, 2214, 2458, 2459, 2496
- COVID-19 impact, on stock returns, 2221
- Covid-19 pandemic, 2560
 - economic uncertainty due to, 2561
- Cox, Ingersoll, and Ross (CIR) model, 825
- Cox model, 1926
- Cox proportional hazard model, 1929, 1930, 1935
- Cox proportional model, 1929
- CPT decision weights, 854, 855, 861, 865
- Credit derivatives, 573
 - asset swaps, 574–575
 - basket default swaps, 581–583
 - convertible bonds, 583
 - credit spread options, 579–580
 - default swaps, 575–577
 - principal protection, 579
 - total return swaps, 577–578
- Credit enhancement, 1040–1042
- Creditors, 970
 - voluntary liquidation, 777
- Credit quality hypothesis, 704
- Credit rationing, 1124–1131, 1133–1135
- Credit risk, 573, 574, 583, 708, 2252
- Credit risk modeling
 - barrier structural model, 1728
 - binomial formulation, 1728
 - calibration, 1747
 - debt prices, 1728
 - default time, 1730
 - derivation of (18), 1752
 - derivation of (24), 1756, 1757, 1759
 - Girsonav transformation, 1754
 - implementation of (24), 1759–1761
 - interest rate, 1729
 - interest rate process, 1753
 - multi-period analysis, 1742, 1743, 1745, 1747
 - multi-variate distributions, 1728
 - random equity and random interest rate, 1755
 - risk neutral measure Q , 1729
 - structural models, 1728
 - “true structural” model, 1728
 - two-period example, 1739–1742
- Credit spread(s), 575, 579, 580, 583, 1294, 1304–1306, 1311, 1314
 - options, 579–580
- Credit valuation model, 1494
- Creditworthiness, 804
- Creditworthy Asian corporations, 1043
- Critical value, 898, 909, 916, 918, 931–933
- Cross-border business, 2458
- Cross-border mergers and acquisitions
 - barriers, 1058, 1060, 1061, 1063, 1065, 1066, 1070
 - bond yields, 1070
 - capital budgeting analysis, 1068

- cash flows, 1069
 - Europe, 1058
 - European firms, 1058
 - European market, 1058
 - European Union, 1058
 - exchange rate, 1070
 - FDI, 1059
 - foreign acquisition, 1068, 1069
 - foreign cash flows, 1069
 - foreign firm, 1058
 - foreign takeovers in the European Union, 1058
 - foreign value, 1069
 - international direct investment, 1059
 - macroeconomic factors, 1059–1062
 - multinational corporation, 1058
 - salvage value, 1069
 - studies, 1059
 - takeovers, 1058
 - tariffs and nontariff barriers, 1058
 - United States, 1059
 - Crossing networks, 995
 - Cross-market volatility transmission effects, 1424
 - Cross-sectional momentum (XSMOM) strategies, 1111, 1113
 - Cross-sectional tests, 1468
 - generalized least squares estimation, 1479
 - risk premia estimation, 1477–1479
 - Cross-section dependence, 2217, 2219
 - Cross-validation, 2180, 2189, 2190, 2197, 2199
 - Crown-jewel lock-up, 2479
 - CRRA
 - and information value, 2640–2641
 - simple portfolio problem, 2645–2651
 - Cryptocurrency, 2178–2180, 2186
 - CTEF model, 2283
 - Cube root, 1939
 - gain-loss functions, 1940
 - Cube root utility function, 1939–1941, 1958
 - analysis of, 1951
 - initial analysis, 1940–1944
 - literature review, 1944–1949
 - Cumulative distribution function, 1990
 - Cumulative hazard function, 1921
 - Cumulative normal distribution function, 1295
 - Cumulative prospect theory (CPT), 853–855, 864, 865, 871–874
 - Cumulative standard normal distribution function, 1300, 1303, 1311
 - Cumulative sum of squares test (CUSUM), 2219
 - Currency, 946, 947
 - crisis, 1179
 - risk, 556, 561, 570, 571
 - Current account deficit, 976, 978, 981
 - Current earnings, 1600, 1612, 1619
 - Current EPS, 1601, 1604–1608
 - Current forward exchange rate, 586
- D**
- Darby's method, 1604–1605
 - individual regression results, 1611
 - partial adjustment coefficient and long-term payout ratio, 1611–1612
 - permanent EPS, 1610
 - pooled regression results, 1615
 - results from pooled regression, 1613
 - statistical information of current and permanent EPS, 1612
 - Data, 1782
 - analytics, 2107
 - bank variables, 1783
 - deposit rates, 1783
 - intensive IT systems, 1877–1879
 - mining, 1766–1768, 1771, 1773–1777
 - natural disasters, 1782
 - privacy, 1881
 - Data envelopment analysis (DEA), 1279
 - Data generating process (DGP), 1800
 - Daubechies Least Asymmetric wavelet transformation, 2582
 - Days' sales in payables (DSP), 1677
 - Day-trading centers, 753, 755
 - DCC-GARCH model, 2566
 - DCF-based methods, 2657
 - Deal certainty, 2478, 2479
 - Deal protection, 2478, 2479
 - Debt, 793
 - financing, 795
 - maturity, 1311, 1312
 - overhang problem, 967
 - structure, 815
 - Debt-to-equity ratio, 793, 2345, 2357
 - Decimalization, 720
 - Decimal trading in the U.S. stock markets
 - bid-ask spreads, 720, 721
 - clustering, 721
 - decimalization, 720
 - equity and exchange-traded options markets, 720
 - equity markets, 719
 - fractional trading, 719
 - individual investors, 719

- Decimal trading in the U.S. stock markets
(*cont.*)
market-making frictions, 721
market quality issues, NYSE post-decimalization, 720
NASDAQ market, 720
NYSE, 720
price improvement, 720
price rounding, 721
return volatility, 720
rounding, 721
serial correlation of price changes, 721
“total” return volatility, 720
trading prices, 719
transaction costs, 721
volatility, 721
- Decision-tree, 2658
- Decision weights, 851, 861, 871–872
- Deep learning (DL), 2179, 2186, 2199, 2209
- Default, 1642, 1644, 1645, 1648, 1650, 1651, 1653, 1654
cost-adjusted mortgage value, 1167
probability, 708, 1298–1300, 1303, 1309, 1310, 1313, 1731, 1735, 1738, 1740, 1743, 1745, 1747
swaps, 575–577
- Default risk, 575, 577, 579, 580
premium, 1296, 1304
- Defined benefit (DB) plans, 2042, 2058
DB firms, 2042
maximum annual retirement benefit, 2042
nontax advantages, 2048
pension coverage, 2043
pension freeze/termination, 2043
potential impact, 2046
- Defined contribution (DC) plans, 2042, 2058
employer’s contribution, 2042
- Defuzzify, 2676
- Dekker method, 916
- Delaware, 730, 732–736
legislature, 2480
responsive corporate code, 738
- Delaware General Corporation Law (DGCL), 2474, 2483
- Deloitte survey, 2495
- Demand shocks, 1138
- Department of Justice (DOJ), 2482
- Dependent variable, 815
- Deposit insurance, 512, 524, 526
banking crisis, 525
benefits, 513, 514
cost, 514, 515
design features, 515, 516, 524
- Depositor discipline, 512, 515
- Deposit rate, 1127, 1780–1782
- Descriptive statistics, 803, 1422–1423
characteristics of sample firms, 2433–2434
- Devaluation trap, 1169, 1175
- DFC approach, 2472
- Difference-in-differences (DID), 794, 809, 1780–1782, 1784–1788, 1792, 1796, 1797
effect, 810
impact, 811
- Director liability, 733, 734, 736–738
- Directors’ duties, 1212
- Dirty surplus flows (DSF), 1855
- Discounted cash flow (DCF), 2124, 2487, 2657
- Discount rate (DR), 631, 1116, 2489
- Discrete state space, 953
- Discrete-time model, 1651
- Discrete-time stationary model, 1649
- Discrete wavelet transform (DWT), 2580
- Discriminant analysis, 778–781
- Diskless Remote Boot Linux (DRBL), 1890–1892
- Dispersed information, 946
- Disposition effect, 1944
- Distress, 774, 776, 782
- Distribution function, 1584, 1586
- Distribution skewness, 1914
- Diversifiable risk, 1401–1403, 1406
- Diversification, 965, 1060
hypothesis, 879, 885
- Diversified mutual fund, 1150
- Divestiture/spin-off process, 2472
- Dividend behavior model, 1601, 1607–1620
- Dividend flexibility hypothesis, 2410, 2413
- Dividend growth model (DGM), 2490
- Dividend-like yield, 902
CAPM method, 910
cost of carry model, 910
sensitivity analysis, 920
- Dividend payment behavior model, 1600, 1601, 1603, 1606, 1613
- Dividend policy model, 2414–2418
- Dodd Frank Act, 1261, 1270
- Domestic currency, 946, 947
- Domestic economy, 946
- Domestic savings, 980
- Double-barrier options, 1652, 1653
- Dow Jones Index, 1991
- DPS, 1625–1639
- Dragon Bonds, 1037
- DRAM chipmaker, 926, 927, 930–932
- DRAM foundry, 897, 898, 925, 928, 930–933

- Drezner's method, 913
 - Dual corporate system, 1191
 - Dual fiduciary duties, 2480
 - Dual legal system, 2482
 - Dual patents system, 793
 - Duffie-Singleton model, 1732, 1733, 1739, 1740, 1747
 - Durable goods, 618
 - Duration analysis, 682
 - complication, 685
 - convexity, 686
 - corporate bonds, 694, 695
 - effectiveness, 692, 693
 - foreign-bond, 689
 - formulation/interpretation, 682, 684
 - future markets, 693, 694
 - immunization, 687, 688
 - market risk, 686
 - price decline, 685
 - price volatility, 684
 - Duration gap, 696, 697
 - applications, 698
 - inflation, 699
 - negative, 697
 - nonfinancial corporations, 699
 - positive, 697
 - Durbin-Wu-Hausman (DWH), 1801, 1813–1816
 - test, 804
 - Dynamic conditional correlation (DCC) model, 2558–2560, 2565, 2569
 - Dynamic economy, 555
 - Dynamic hedging, 1412, 1413, 1428
 - Dynamic model, 2410
 - Dynamic proportion portfolio insurance (DPPI), 2198
- E**
- Earnings management, 2042, 2045, 2046, 2058, 2063
 - accruals manipulation, 2064–2066
 - consequences and determinants, 2068–2073
 - different types, 2080
 - discern and measure, 2074–2080
 - real activities manipulation, 2066–2068
 - trade-off between accruals manipulation, 2072–2074
 - varied managerial incentives for, 2062–2064
 - Earnings response coefficient (ERC), 2551
 - ECN Rule, 725
 - Econometrics, 2178, 2179, 2203
 - Economic capital, 2130–2132
 - Economic factors, 2657
 - Economic models, 1554
 - Economic policy uncertainty (EPU), 587, 595, 596, 598, 599
 - Economic theory, 884
 - Economies of scale, 963, 964
 - Edogenous termination, 1163
 - Effective exchange rates, 1140, 1147
 - Effective post-merger integration, 2459
 - Efficiency frontier curve, 1152
 - Efficiency theories, 879–881
 - Elango model, 1286, 1287
 - Electronic call auction, 998–1000
 - Electronic communication network (ECN), 723, 754, 1100
 - Electronic trading, 1010, 1015
 - Empirical price illusion, 1947
 - Employee Retirement Income Security Act (ERISA), 2043, 2058
 - Employees' remuneration, 887
 - Employer's defined pension obligation (DBO/PBO), 2044, 2045, 2047
 - Employment Act of 1946, 606
 - End-of-the-month spot exchange rates, 587
 - Endogeneity problem, 804
 - Endogenous default model
 - Geske (G) model, 1306–1310
 - Leland/Leland and Toft (LT), 1310–1314
 - Endogenous uncertainty, 1361
 - Energy valuation model, 1497
 - Enhanced-index-tracking (EIT) portfolio strategy, 2291
 - Ensemble, 2180, 2197, 2199–2206, 2208
 - Enterprise value (EV), 2486
 - Enterprise value to earnings before interest and tax (EV/EBIT), 2486
 - Entropy, and information value, 2639–2640
 - Epstein–Zin model, 619
 - Equilibrium asset pricing models, 614
 - Equilibrium credit rationing
 - financial intermediary, 2230
 - firm's investment projects, 2230
 - institutional and legal restrictions, 2230
 - interest rates, 2231
 - literature, 2234, 2235
 - number of loans, 2233
 - size of loans, 2233
 - types, 2230
 - vector-error correction model, 2231
 - Equilibrium models, 825–827
 - Equity financing, 889
 - Equity market-makers, 1534

- Equity premium, 587, 593, 595, 596, 598
- Equity premium puzzle
 - adjusted set of measurements, 1358, 1359
 - borrowing constraints, 1355, 1356
 - calculations and results analysis, 1366–1368
 - challenges, 1346
 - countries, 1350, 1351
 - data sources, 1346
 - definition, 1362
 - disaster state and survival bias, 1354, 1355
 - equity investments, 1346
 - financial economics, 1346
 - generalizations, 1346
 - habit formation, 1352–1354
 - idiosyncratic and uninsurable risk, 1354
 - incomplete market, 1357, 1358
 - interpretations, 1362
 - investors planning, 1362
 - literature, 1351
 - Lucas' Asset Pricing Model (CCAPM), 1347–1350
 - Merton's model, 1360, 1361
 - the model, 1365, 1366
 - portfolio optimization, 1369
 - rational beliefs, 1361
 - risk measure, 1364, 1365
 - stable distributions, 1362–1364
 - standard economic models, 1346
 - standard preferences, 1347–1350
 - statistical data analysis, 1359, 1360
 - taxes, 1356, 1357
 - time and state preferences, 1351, 1352
 - transaction costs, 1356
- Equity prices, 1181–1183
- Equity risk premium differential hypothesis, 591, 592
- Equity risk premium hypothesis, 592
- Equity risk premium puzzle, 860, 861
- Equity-valuation models, 2079–2087
- Error correction model (ECM), 1716, 1724
- Errors-in-variables (EIV), 2370
 - Bayesian approach, 2385–2386
 - bivariate normal case, 2371–2372
 - classical estimation method, 2373–2376
 - finance research applications, 2386–2405
 - grouping method, 2376–2378
 - instrumental variable method, 2378–2380
 - LISREL and MIMIC methods, 2383–2385
 - mathematical method, 2380–2382
 - maximum likelihood method, 2382–2383
 - multivariate case, 2372
- Escrow account, 545
- Estimation errors, 750–751
- E* Trade, 754, 756
- Euler equations, 559–560, 565
- Euler-Lagrange condition, 2419
- Euro-currency deposit rates, 587
- Euro-dollar futures options, 820
- Europe, 1226
- Europe, Australia, and Far East Index (EAFE), 1276
- European Banking Authority (EBA), 1984
- European Central Bank, 2235
- European Economic Community, 1226
- European Exchange Rate Mechanism (ERM), 1252
- European Monetary System (EMS), 1078
- Evaluating convertible bond
 - calculus models, 1565
 - callable vs. non-callable, 1564
 - conversion policy, 1565
 - decision variables, 1565
 - dynamic stochastic models, 1563, 1565
 - dynamic stochastic stock price, 1573, 1574
 - firms, 1564, 1565
 - firm's market values, 1563, 1565, 1566
 - fixed claims, 1564
 - geometric Brownian motion, 1563, 1565
 - graphical model, 1566, 1567
 - hazard rate, 1566
 - instruments, 1564
 - interest rate, 1563
 - investor protection, 1578
 - non-structural models, 1565
 - security design, 1578
 - shareholders, 1577
 - spread and reduced-form approaches, 1573, 1574
 - state variables, 1565
 - static stochastic models, 1563, 1565
 - static stochastic stock, 1567–1570
 - straight bond prices, 1567–1570
 - structural models, 1565, 1566, 1570–1573
 - valuation methods, 1566
 - valuation model, 1563
- Event studies, 1179, 1180, 1183
- EVTGARCH, 2179
- Ex-ante-funded scheme, 524, 525
- Ex ante* real rate, 603
- Ex-ante* synergy, 1064, 1065
- Exceedance, 2184, 2193, 2194, 2200, 2201
 - probability, 2178, 2182, 2184, 2191, 2195
- Excess Spread Rule (ESR), 725
- Excess Value Created (EVC), 1860
- Exchange Act Release No. 14416, 1098

- Exchange rate(s), 1059, 1060
 - policy signal, 1140
 - response function, 1139
 - risk, 570, 689
 - Exchange rate market volatility (EMV), 586, 588, 591
 - Exchange traded funds (ETFs), 605, 1265
 - Executive compensation, 2072, 2140
 - Executive Directors, 1194
 - Executive Meeting of East Asia and Pacific (EMEAP) Central Banks, 1038
 - Exercise prices, 928
 - Existing rules, 1671, 1675
 - Exogenous default model
 - Black and Cox (BC), 1299–1302
 - Collin-Dufresne and Goldstein (CDG), 1304, 1305
 - Longstaff and Schwartz (LS) model, 1302
 - Exogenous prepayment, 1163
 - Expanded NPV, 898
 - Expectation hypothesis, 1911, 1914
 - Expected asset returns, 619
 - Expected rate of return (ERR), 2046
 - Expected return, 744
 - Expected shortfall (ES), 1443, 1447, 1883
 - Expected termination probability, 1170
 - Expected utility, 855
 - Expected utility theory (EUT), 873
 - Experimental studies in finance
 - arbitrage models, 867–868
 - asset allocation, 866
 - average subject behavior, 873
 - CAPM, 867, 873
 - cash-flows integration, 873
 - and decision weights, 871–872
 - diversification, 866, 867
 - economics, 858
 - equity risk premium puzzle, 860, 861
 - financial theory, 874
 - FSD dominance, 870
 - investment horizon, 866
 - mean–variance dominance, 870
 - M–V rules, 869
 - portfolio diversification, 858–860
 - portfolios and mutual funds, 869–871
 - PT, 869
 - random walk, 858–860
 - rational models, 874
 - SD rules, 868, 869
 - shape of preference, 861–866
 - Exponential GARCH (EGARCH) model, 2559
 - Exponential smooth transition model (ESTR), 1689
 - Exponential utility function, 2416
 - Ex post* real rates, 607
 - Ex-post scheme, 525
 - Exposure priority, 1095
 - Extant literature, 1111
 - Extended merton model, 1733–1737
 - Extended simultaneous volatility (ESVL) framework, 1419
 - External auditors, 1191, 1198, 1199, 1210
 - External capital market, 2118, 2119
 - relative dependence, 2122
 - External controls, 1210
 - functions, 1210
 - independent auditors, 1210
 - External equity financing, 889
 - External investment through M&As, 2460, 2462
 - Extra-effort aversion, 1588
 - Extra-ordinary risk, 1984
 - Extreme value theory (EVT), 2179
- F**
- Fabozzi, F., 1398
 - Factor analysis, 696
 - Factor analytic/principal components methods, 622
 - Factor models, 622–624
 - Failure prediction, 788
 - Fair Market Value (FMV), 2486
 - Fairness opinions, 2460
 - Fama, E.F., 1398
 - Fama-French model, 1990
 - Fama model, 1284, 1285
 - Family-owned firm, 1194
 - Fannie Mae, 844
 - Feasible generalized least squares (FGLS), 1807, 1815
 - Feasible General Least Squares Estimation Procedure (FGLS), 783
 - Federal agencies, 2483
 - Federal Deposit Insurance Corporation (FDIC), 1984, 1995
 - Federal Funds rate, 1080
 - Federal Housing Administration (FHA), 844, 845
 - Federal Reserve Economic Database (FRED), 2236
 - Federal Reserve's discount window policy, 1179
 - Federal Reserve System, 514
 - Federal Trade Commission (FTC), 2482
 - Fed model, 757

- Female directors
 - positive, 1205
 - trend, 1205
- Ferruz, L., 1404
- Fiduciary duties, 2480
- Fiduciary out, 2479
- Finance leases, 2027, 2028, 2037
- Finance services, information technology in,
 - see* Information technology (IT), in financial services
- Finance theory, 970
- Financial accounting, 659
- Financial Accounting Standards Board (FASB), 1659, 2020, 2025, 2046, 2047, 2493
 - financing activities, 1660, 1661
 - investing activities, 1660
 - operating activities, 1660
- Financial arrangements, 792
- Financial assets, 613, 614, 1132
- Financial control systems, 1559
- Financial crises, 774, 1177
- Financial crisis (2007-2008), 2214
- Financial decisions, 1583
- Financial disclosures, 1218
 - obfuscation, 1218
 - opaque, 1218
 - timely, 1219
 - transparency, 1218, 1219
- Financial distress, 2070
- Financial engineering, 542
- Financial firms, 2124
 - economic capital, 2130–2132
 - performance measures, 2126–2127
 - regulatory capital, 2127–2130
- Financial flexibility, 815
- Financial holding companies, 531
- Financial Information Grid (FinGrid), 1875–1876
- Financial instability, 1532
- Financial institution management, 686, 696
- Financial institutions, 969, 1041
- Financial intermediaries, 1043
- Financially constrained, 2070
- Financially distressed firms, 2142, 2145, 2148, 2154, 2156, 2158, 2162, 2169, 2170, 2172
- Financial market, 755, 1531, 1537
 - development, 792
 - equilibrium, 614
- Financial misstatements, 2107
- Financial modernization, 536
- Financial modernization Act, 530
- Financial panel data models
 - advantages, 1825
 - balanced vs. unbalanced panel data, 1813
 - clustering models, 1827
 - contemporaneous, 1801
 - cross-sectional/time-series data, 1800
 - cross-sectional effects, 1821
 - cross-sectional random effect, 1823
 - Durbin-Wu-Hausman (DWH) tests, 1813–1816
 - dynamic panel GMM estimator, 1822
 - endogeneity, 1801, 1802
 - error component, 1826
 - estimation, 1821
 - exogeneity, 1801, 1802
 - experimental data, 1817, 1818
 - explanatory variables, 1826
 - FE vs. RE specification, 1826
 - financial analyses, 1821
 - firm and time effects, 1824
 - firm effects, 1822
 - first-difference (FD) estimator, 1827
 - fixed effects models, 1800, 1807–1810
 - fixed time effects, 1824
 - hypothesis tests, 1800
 - linear panel data models, 1802–1806
 - multicollinearity, 1826
 - omitted variables, 1799
 - panel data estimators, 1819
 - parameters, 1818
 - pooled LS, 1821
 - pooled LS estimation, 1802–1806
 - random effects estimators, 1826
 - random effects models, 1810–1813
 - statistical properties, 1825
 - strict exogeneity, 1801, 1802, 1823
 - test strict exogeneity, 1816
 - time-invariant regressors, 1826
 - two-step estimation process, 1818
 - unobservable heterogeneity, 1807
 - unobserved heterogeneity, 1825
 - Wald test, 1823
- Financial planning and forecasting
 - algebraic simultaneous equations, 2345–2357
 - characteristics, 2344
 - Francis and Rowell model, 2358–2362
 - inputs, models and outputs for, 2343
 - percentage sales method, 2344–2345
 - sensitivity analysis, 2357–2358
- Financial ratio(s), 1661
 - analysis, 778–781
- Financial regulators, 1048, 1052

- Financial reporting incentives, 2024, 2025
- Financial risk, 2070
- Financial shock, 2123
- Financial simulation
 - option pricing, 1882–1883
 - VaR, market risk measurement based on, 1883–1884
- Financial stability, 774
- Financial system, 513, 1259
- Financing innovation, 795
- Financing policy(ies), 2140, 2442
- Financing strategies, 1642, 1651, 1655
- FINPLAN model, 2345
- FinPlan program, excel procedure in, 2364–2367
- Fire-sale, 513
- Firm cash flows, 1586
- Firm characteristics, 1117
- Firm performance, 1198, 1213, 1215
 - financial expertise, 1217
 - managers/shareholders interest, 1218
 - outside directors, 1216, 1217
- Firm's capital budgeting process, 2122
- Firm's internal capital budgeting, 2122
- Firm's investment, 2140
- Firm size, 1686, 2069
- Firm's market values, 1571, 1576
- First Asian bond fund (ABF-1), 1038
- First degree Stochastic Dominance (FSD), 852–855, 861, 863, 866, 867, 869–872
- First-in first-out (FIFO), 2493
- First-order Taylor approximation, 567
- Fisher equation, 1914
- Fisher–Weil duration, 695
- Fisher–Weil process, 692
- Fixed effects (FE), 1801, 1803, 1807–1810, 1812–1824
 - models, 2436
 - regressions, empirical structural change point, 2437, 2440
- Fixed exchange rate model, 1125
- Fixed exchange rate system, 1132
- Fixed income securities, 682, 683, 686, 687, 697
- Flexibility gain, 1552
- Flexible decision-making, 899
- Flexible immunization approach, 688
- Floating-rate coupon bonds, 1304
- Floating-rate payments, 1303
- Flow of funds
 - accounting perspective vs. financial perspective, 658–659
 - accounts receivable, 657–658
 - cash, 656
 - cash flow cycle, 662
 - integrated cash management system, 661–662
 - inventories, 658
 - marketable securities, investing in, 660–661
 - marketable securities, 656
 - reasons for holding cash, 659–660
- Flows to-equity (FTE), 2488, 2491
- Fong–Vasicek measure, 693
- Forbes' Billionaires List, 1254
- Forecasting model, 1946
- Forecast performance evaluation, 2574–2576
- Forecast tracking error, 2289
- Foregoing regression analysis, 2079
- Foreign acquisition, 1068, 1069
- Foreign-bond portfolio, 689
- Foreign capital inflow, 974, 976, 978, 981
- Foreign currency, 946, 947
- Foreign direct investment (FDI), 1058
- Foreign-exchange rate, 690
- Foreign exchange reserves, 974–976, 978–980
- Foreign exchange risk premium (FXRP), 586
- Foreign exposed banks, 1182
- Foreign firms, 1067
- Foreign investors, 1047
- Foreign portfolio investment, 980
- Fortran, 911, 917, 918
- Forward exchange rate risk premium, 589, 598
- Forward premium hypothesis, 588, 589
- Forward rates, 1907, 1911–1912
- Four factor model, 2509
- Fourier-cosine expansion, 1987
- Fourier expansion, 915
- Fourier index, 915
- Fourier series-based method, 1985, 1989, 2003
- Fractional trading, 719
- Fragmented market, 1101, 1102, 1107
- Fragmented markets, 1014
- Francis and Rowell model
 - econometrical advantage, 2362
 - equations for, 2361
 - objectives, 2358
 - sectors, 2362
 - variables for, 2358–2360
- Franklin Templeton mutual fund, 1279
- Frauds, 2107
- Freddie Mac, 844
- Free cash flow, 659, 883, 1067, 2488, 2490
 - hypothesis, 879, 883–884, 2413
- French, K.R., 1398
- Frictionless markets, 1351

Frictions, 958
 Friedman and Savage utility function, 857
 Friedman test, 2630
 Friendly merger, 2466
 Friendly transaction, 2466
 Frog-in-the-pan (FIP), 1114
 Full information maximum likelihood (FIML)
 estimation, 1413, 1425–1426
 Full-service brokerage, 755
 Fundamental benefits of publicly traded
 corporation
 diversification, 965
 economies of scale, 964–965
 liquidity, 965
 Fundamental costs of publicly traded
 corporation
 information asymmetry, 967
 principal–agent conflicts of interest,
 966–967
 Fundamental Theorem of Asset Pricing, 952
 Funding, 1966, 1969
 Funding options
 pre-funded coupon bonds, 543, 544
 zero-coupon bonds, 542, 543
 Fund performance, 669, 670, 672
 Future returns, 2504, 2525
 and abnormal sales growth, 2508
 Future spot rate, 586
 Fuzzy capital budgeting approaches, 2656,
 2674
 Fuzzy cash flows, 2656, 2658, 2666, 2668,
 2671, 2674
 Fuzzy decision, 2676
 Fuzzy logic, 2676
 Fuzzy numbers, 2656, 2663, 2664, 2674
 Fuzzy real option valuation (FROV),
 2665–2667, 2671, 2674
 posterior probability, 2660, 2662, 2667
 prior probability, 2660
 probability, 2659
 real option, 2658, 2659
 ROV, 2659
 thinking in real options, 2659
 traditional methods, 2659
 Fuzzy sets, 2675, 2676

G

GARCH(1,1)-BEKK model, 1724
 GARCH, 2178–2181, 2186, 2192, 2194–2196,
 2199, 2209
 GARCH model, 1088
 Garrett and Priestley's method, 1605

Gauss–Hermite jump process, 1078, 1079
 Gaussian models, 1734
 Gaussian shock, 1080
 Gauss-Markov Theorem, 1808
 Gauss quadrature method, 911, 913, 932
 Generalized autoregressive conditional
 heteroskedasticity (GARCH), 561
 bivariate, 1713
 conditional variance and covariance,
 1712, 1713
 conventional regression model, 1716
 development, 1712
 error correction estimates, 1717
 heteroscedastic errors, 1716
 model, 2558
 summary statistics, 1715
 unit roots and residual cointegration
 tests, 1717
 Generalized expected utility (GEU), 1351
 Generalized functional form (GFF),
 1687, 1688
 Generalized least squares (GLS), 1479
 Generalized method of moments (GMM),
 561, 674–677, 1172, 1173,
 2397, 2615
 advantages, 1470
 estimator, 1471–1472, 2379
 framework, 1358
 Hansen-Jagannathan-distance, 1475–1477
 moment conditions for asset pricing tests,
 1472–1474
 Generalized Pareto distribution (GPD), 2181,
 2194, 2195
 Generalized semivariance (GSV), 1705
 Generally accepted accounting principles
 (GAAP), 2020, 2493, 2504
 Genetic algorithms, 1031
 Genetic programming, 1031
 Geometric Brownian Motion, 1295, 1299,
 1300, 1302, 1305, 1306, 1310,
 1565, 1571
 Geometric jump diffusion model, 1082
 Geopolitical indices, 587
 Geopolitical risk (GPR), 588, 596, 598, 599
 Geopolitics, 1049
 George Soros' Quantum Fund, 1253
 Geske (G) model, 1306–1310
 Geske-Johnson model, 1736–1739, 1741, 1743,
 1747, 1756, 1759
 Ginnie Mae, 844
 Ginnie Mae HECM Pooling (HMBS), 1379
 cashflows, 1379–1380
 trust assets, 1388–1391

- Ginnie Mae Reverse Mortgage, 1382–1384
Girsanov's theorem, 1170
Glassman–Hassett model, 757
Glass-Steagall Act (GSA), 530
Global Analyst Forecasting Efficiency Model (GAFEM), 2284
Global banks, 1236–1238
Global Covid-19 health crisis, 2462
Global Economic Policy Uncertainty (GEPU), 587
Global empirical analysis, 2599, 2600, 2622
Global finance
 ample liquidity, 1044
 asset management, 1045
 emerging market currency bonds, 1045
Global financial crisis, 2620
Globalization, 539
Global mutual funds
 annual return risk, 1275
 Beta coefficient, 1275
 Chang, Hung and Lee Model, 1287
 Choi model, 1286
 decision-making structure, 1277
 Delphi method, 1281
 Elango model, 1286, 1287
 equilibrium asset pricing model, 1278
 equilibrium return, 1275
 Fama model, 1284, 1285
 financial instruments, 1274
 financial performance, 1281
 fund risk, 1276
 geographic asset, 1278
 identification of risk, 1278
 Indian mutual funds, 1278
 international mutual funds, 1282
 investment managers, 1275
 investment period, 1280
 Italian stock exchange, 1279
 Jensen model, 1284
 Malaysia, 1278
 MM approach, 1287–1289
 mutual fund industry, 1281
 NAVs, 1276
 pension and investment fund performance, 1276
 performance evaluation, 1274
 risk-adjusted future performance, 1277
 Statman model, 1285
 stock returns, 1277
 Treynor and Mazuy model, 1285
 U.S. market, 1277
 67 US mutual funds, 1277
Global portfolio investment, 980
GNR 2011-H10
 assumptions for, 1382–1384
 trust assets, 1388–1391
Golden parachutes, 1212, 2477, 2478
Goldilocks scenario, 1139
Gómez-Bezares, F., 1402, 1404, 1408
Gómez-Bezares, F.R., 1408
Good-deal option pricing bounds, 673
Goods market approach, 946
Goodwill, 888
“Go-shop” provisions, 2479
Government guarantee, 512
Gradient Boosted Regression Tree (GBRT), 2107
Gramm-Leach-Bliley Act (GLBA), 530
 bank regulations, 537
 equal treatment, 534
 financial holding companies, 531
 functional regulation, 533
 implications, 537–539
 insurance, 530
 national bank financial subsidiaries, 531
 potential benefits, banks, 535
 potential risk elements, customers, 535–537
Graphical method, 1565
Graphical model, 1566, 1567
Graphic process unit (GPU), 1880
Great depression, 1178
Great recession, 1241
Grid computing, 1871, 1872, 1874, 1887, 1894, 1896
Gross Domestic Product (GDP), 512
Grouping method, 2376–2378
 EIV problem, 2376
Growth-style benchmark index, 985
Guideline companies, 2492
- H**
Habit formation, 1352–1354
Habit persistence, 618, 619
Hansen and Jagannathan distance (HJD), 677, 1475–1477
Hart-Scott-Rodino Act (HSR), 2482
Hart-Scott-Rodino Antitrust Improvements Act (HSR/HSR Act), 2482
Hazard models, 787
 coefficient, 1934, 1935
 empirical application, 1931, 1933
 function, 1921
 hazard models, 1923, 1928
 Kaplan-Meier estimator, 1922
 logistic regression, 1925–1928

- Hazard models (*cont.*)
 parametric, 1923, 1924
 semi-parametric, 1925
 statistical methods, 1919
- Hazard regression, 1920
- Hedge Fund Research, Inc. (HFRI), 1256
- Hedge Fund Research (HFR), 1266
- Hedge funds
 arbitrage, 1261, 1263
 asset growth, 1254
 backfill bias, 1254
 bear markets, 1254
 definition, 1250
 Equity and Bond Summary Statistics, 1256
 fees, 1265, 1266
 financial instruments, 1250
 financial markets, 1250
 Ford and General Motors, 1263, 1264
 fund performance, 1255
 history, 1250, 1251
 increased liquidity, 1267
 increased regulation, 1269, 1270
 influence, 1250
 lennar corporation, 1263
 managers, 1251–1253
 Motorola, 1265
 organization, 1259–1261
 payment, 1266, 1267
 public vehicles, 1268, 1269
 regulation, 1259–1261
 sleepy cottage industry, 1253
 strategies, 1261
 transparency and growth, 1267, 1268
 trends, 1265
 types, 1256–1259
 U.S. Equity and Bond Performance, 1255
 Wall Street, 1253
- Hedge ratios
 academicians, 1704
 approaches, 1706
 cointegrated heteroscedastic methods, 1718
 cointegration method, 1706
 conditional distributions, 1705
 conditional heteroscedastic method, 1706
 conventional assumptions, 1718
 econometric methodologies, 1718
 financial economic theories, 1718
 future contract, 1704
 investments, 1705, 1706
 lower partial moments, 1705
 margin, 1704
 minimum value-at-risk, 1711
 multi-period model, 1705
 MV, 1704, 1707, 1708
 optimal, 1704, 1705, 1707, 1717
 optimum mean-variance, 1708
 practitioners, 1704
 random coefficient method, 1706
 return distributions, 1705
 risk management, 1704
 R language, 1723, 1724
 S&P500 index, 1719–1723
 sharpe, 1709, 1710
 statistical properties, 1718
 techniques, 1706
 utility functions, 1705, 1718
- Hedging, 1020–1021
 interest rate risk, 686, 698
 risk, 556, 561, 570, 571
- Helaba, 1984
- Herfindahl-Hirschman Index (HHI), 2482
- Heterogeneity, 2332
- Heterogeneous mortgages, 1165
- Heteroscedasticity, 815
- Heteroskedasticity and autocorrelation
 consistent (HAC), 588
- High- and low- innovative firms, 803
- Higher-quality innovations, 795
- High-frequency serial dependence, 633
- High-quality accruals, 2077
- High-quality liquid assets (HQLA), 2254
- h-index, 1966
- Hi-tech value, 902–904
- Ho, L.-C., 1398
- Hodges, C.W., 1398
- Holding cash, 659–660
- Holding free cash flow, 883
- Holding period return (HPR), 688
- Ho–Lee model, 827, 829, 832, 1488
- Home equity conversion mortgages (HECMs)
 cash flows and prepayments, 1379–1380
 fees and caps, 1377–1378
 Ginnie Mae HECM pooling, 1379
 GNR 2011-H10, 1382–1384
 vs. HELOC, 1375–1376
 pool characteristics, 1379
 PPC curve, 1378–1382
 prepayment curve, 1378–1382
 principal limit and maximum claim amount, 1377
 qualifications and payment plans, 1376–1377
 repayment speeds for, 1394
- Home equity line of credit (HELOC), 1375–1376
- Homogenous transaction cost model, 1160

- Homoscedasticity, 780
- Hostile bid, 2467
- Hostile bidders, 2475
- Hostile bidders tactics
 - bear hug, 2466
 - friendly transaction, 2466
 - proxy fight, 2468
 - SEC, 2467
 - tender offer, 2468
- Hostile takeovers, 2465
- Hostile tender offer, 2468
- Hostile transaction, 2469
- Housing devaluation effect, 1175
- Housing price process, 1157, 1160, 1161, 1171
- Housing prices, 1160–1163, 1166, 1167, 1169, 1170, 1175
- H-REMIC, 1380
- Huarong Asset Management, 1051
- Hubris hypothesis, 882
- Hull–White model, 832
- Hybrid market, 992, 1016
- Hypothesis developments
 - asymmetric timeliness, 1834
 - cash flow vs. accruals, 1835
 - dirty surplus flows, 1836, 1837
 - line items, 1835, 1836
- I**
- Identity weighting matrix, 677
- IFRS 16, 2037, 2038
- IMF bailouts in international debt crises
 - Asian crisis in 1997, 1181
 - future research, 1183
 - ILSA, 1180
 - informational efficiency, 1182
 - LDC loan, 1179
 - loan commitments, 1181
 - negative view, 1179
 - non-South Korean exposed bank group, 1182
 - pricing notion versus the rational pricing notion, 1181
 - South Korean exposed bank group, 1182
 - zero coefficients, 1181
- Immediacy, 1008
- Immunization, 682, 686–688, 691, 693–696
 - complication, 690
 - maturity bonds, 691
 - validity, 692
- Impact investing, 1964–1966, 1969, 1970
- Implied default probability, 708, 709
- Implied hedge ratio (IHR), 1418
- Implied standard deviation
 - accuracy, 1520, 1521, 1523, 1524
 - B/S model, 1515
 - Black-Scholes option pricing model, 1512, 1514, 1526
 - Brenner and Subrahmanyam's approach, 1512
 - Chance's formula, 1513
 - constant variance assumption, 1513
 - Corrado and Miller's formula, 1513
 - cumulative normal distribution, 1515
 - different exercise prices, 1518–1520
 - estimation error, 1516
 - finance literature, 1512
 - option pricing model, 1512, 1514
 - roles, 1512
 - single call option, 1516–1518
 - Taylor expansion, 1514, 1527
 - Taylor series expansion, 1513, 1515
- In, F., 1398
- Incentive plan, 2147
- Incentives, 1019, 1023
- Income valuation approach, 2486–2488
- Incorporation, 730, 732, 734
- Incremental profits, 1429–1430
- Independent directors, 1018, 1022–1024, 1194
 - specialization, 1196
- Independent outside directors, 1193
- Indian Stock Market, 1536
- Individualism, 1118
- Individual momentum, 1110
- Individual outside directors, 1204
- Individual *t*-test, 1469
- Inductive reasoning, 1031
- Industrial Groupings in Japan*, 797
- Industry-level job security, 2145, 2152
 - measure, 2169
- Industry momentum, 1110, 1111, 1113
- Inflate earnings, 2077
- Inflationary economy, 1146
- Inflation contingent claims valuation model, 1498–1499
- Inflation costs, 1141
- Inflation information aggregation, 608, 609
- Inflation-protected securities, 605, 606
- Inflation rate, 587, 1143
- Inflation risk, 699
 - premium, 606
- Information, 937–942
 - asymmetry, 964, 965, 967, 2120
 - diffusion, 1118
 - efficiency, 1182
 - hypothesis, 879, 886

- Information (*cont.*)
 risk, 609, 610, 2071
 technology, 999
- Information discreteness (ID), 1115
- Information technology (IT), in financial services, 1870–1871
 artificial intelligent-based high performance computing platform, 1889–1890
 at-home style PC grid, 1893–1894
 big data and MapReduce systems, 1879–1880
 cloud-enabled hybrid systems, for big data and machine learning, 1880–1881
 competitiveness, 1871–1873
 data intensive IT systems, 1877–1879
 DRBL cluster, 1890–1892
 financial simulation, 1882–1884
 FinGrid, 1875–1876
 Monte Carlo simulation, 1884–1887
 PC grids, 1876
 PicsouGrid, 1875
 PRAGMA grid, 1892–1893
 RBNB DataTurbine, 1895
 service integration model, 1876
- Information value
 and CRRA utility, 2640–2641
 and entropy, 2639–2640
 and risk aversion, 2639
- Initial public offering (IPO) price, 898, 903, 908, 917, 918, 924, 925, 931
- Innovation, 813
 activities, 815
 bank debt, 792
 description, 792
 efficiency, 813, 814
 financing, 794
 financing flexibility, 795
 investments, 792
 main bank relationship, 795–797
 nature, 794, 795
 projects, 792
- Insiders, 1192
- Insolvency, 774
- Insolvency problems, 774
- Institute for International Finance, 1245
- Institutional investors, 754, 757, 758, 1196, 1214, 1215, 1268
- Institutional ownership, 1686
- Institutional Shareholder Services (ISS), 2476
- Instrumental principal component analysis (IPCA), 1116
- Instrumental variable method, EIV problem, 2378–2380
- Insurance companies, 682, 698
- Integrated cash management system, 661–662
- Interactive effect, 812
- Interest rate derivatives, 820
- Interest rate models, 1493
 balance sheet and limitations of, 1913
 validity of, 1904–1905
- Interest rate movements, 820
- Interest rate shock, 692
- Interest rate volatility, 822, 823
- Intermarket Trading System (ITS), 1098, 1101, 1102, 1104, 1107
- Intermediaries, 969
- Internal capital budgeting
 agency problem, 2119–2121
 of financial firms, 2124–2132
 firm boundary, 2121, 2122
 issues related to, 2119–2124
 of non-financial firms, 2132
 optimal timing, 2121
- Internal capital market, 2118–2120, 2122
- Internal controls, 1206
 auditing standards, 1209
 historical perspective, 1206, 1207
 SOX, 1207
 Section 30, 1207, 1208
 Section 404, 1208
 Section 404a/404b, 1208
 Section 404b, 1208, 1209
- International Accounting Standards Board (IASB), 2020, 2026, 2037
- International asset pricing model (IAPM), 556, 557, 560, 561
 PPP deviate, 569–570
 without consumption, 567–569
- International bank creditors, 1178, 1183
- International banking
 2008–2009 Financial and European Sovereign Debt Crises, 1234–1236
 2008–2009 financial crisis, 1223
 Basel III policies, 1242–1245
 currency of issue, 1231–1233
 Europe, 1226
 failure of basel II, 1241, 1242
 financial integration, 1223
 global banks, 1236–1238
 history, 1239–1241
 international bonds and notes, 1233, 1234
 international debt markets, 1230, 1231
 international money market instruments, 1231–1233
 Japan, 1224
 policymakers, 1223

- relationship to economic activity, 1227–1230
 - United States (US), 1225
 - International Banking Facilities (IBFs), 1225
 - International debt markets, 1178, 1181, 1182
 - International direct investment, 1057–1059
 - International financial community, 979
 - International financial reporting standards (IFRS), 2020
 - International Lending Supervision Act (ILSA), 1180
 - International merger and acquisition, 1062
 - International momentum, 1118
 - International Monetary Fund (IMF), 512
 - International pure earnings yield factor portfolios, 2307
 - International pure momentum factor portfolios, 2308
 - International rating agencies, 1040, 1043
 - Internet, 754, 757, 758
 - Intermediation process, 2050
 - Intertemporal asset pricing model, 556, 560
 - Intertemporal Capital Asset Pricing Model (ICAPM), 749, 750
 - Intertemporal marginal rate of substitution (IMRS), 614–616, 618, 619
 - Intertemporal risk, 570
 - Intertemporal substitution, 556, 557, 561, 562, 568, 570
 - Intraday momentum, 1112
 - Intra-day price volatility, 1000
 - Intra-state banking deregulation, 792
 - Inventories, 658
 - Investment, 2020, 2031, 2034, 2035
 - Investment companies
 - AXP and CSCO, 1992, 1993
 - parameters of residuals, 1992
 - risk-free assets, 1991, 1992
 - Investment Company Act, 1150
 - Investment Company Act of 1940, 1259
 - Investment Company Act of 1996, 1260
 - Investment decision making, 2657
 - Investment equation, measurement error in, 2403–2405
 - Investment performance, 639, 640, 647
 - Investment policy, 2442–2444
 - Investment project, 896–900, 902–904, 907, 908, 911, 920, 923, 925, 928, 930, 932, 933
 - Investments, 2458
 - Investor arbitrageurs, 1321, 1324, 1333
 - Investor overconfidence, 1583
 - Investor's preferences, 2180
 - IPO firms, 792
 - Irimia-Diéguez, A.I., 1404
 - IS-LM-BP curves, 1133
 - Italian manufacturing SMEs, 794
- J**
- Jackpot payoffs, 1947
 - Japan, 1224, 1776
 - Jarrow-Turnbull model, 1731, 1732, 1740, 1743
 - Jensen alpha, 641, 646, 763–765, 985–987, 1398, 1404, 1406
 - approach, 2509
 - Jensen measures, 987
 - Jensen model, 1284
 - Jensen's inequality, 2637
 - Job insecurity, 2140
 - Job market, 2143
 - Job security, 2144
 - cash pay, 2144
 - and CEO compensation, 2152, 2154, 2156, 2158
 - commitment, 2140
 - under different labor market conditions, 2164, 2165, 2167, 2169
 - under different risk preferences, 2158, 2159
 - exogenous, 2140
 - firm innovation, 2142
 - firm-level measures, 2158
 - in initial pay negotiation, 2140
 - measures, 2144–2146, 2151–2153
 - proxy, 2144
 - and worker's income spans, 2140
 - Job vacancies, 2143
 - Johnson & Johnson's financial statement and share price, 2363–2364
 - Johnson and Johnson Company
 - balance sheet, 1670
 - comparative balance sheets, 1663, 1664
 - consolidated statements of earnings, 1665
 - corporations, 1670
 - distribution network, 1671
 - financial statements, 1662
 - financing activities, 1667–1670
 - income statement, 1663, 1664, 1670
 - indirect method, 1666
 - investing activities, 1667
 - operating activities, 1663, 1665, 1667
 - Joint Energy Development Limited Partnership (JEDI), 1021
 - Joint *F*-tests, 1469–1470

Joint hypothesis problem, 939
 Joint ventures (JVs), 2473
 Jump diffusion model
 affine jump diffusion, 1081–1082
 ARCH/GARCH jump diffusion model, 1087–1089
 autoregressive jump process model, 1083–1084
 Bernoulli jump process, 1077
 conditional jump dynamics model, 1085–1087
 Gauss–Hermite jump process, 1078–1079
 geometric jump diffusion model, 1082
 jumps in interest rates, 1079–1081
 mixed-jump processes, 1075–1076
 Jump diffusion with conditional heteroskedasticity, 1087
 Jump frequencies, 1089
 Jump intensity, 1085
 Jumpstart Our Business Startups Act (JOBS Act), 1260, 1261
 Jump transform, 1082
 Juniper Inc. (JNPR) for 3 February 2003, 1104
 Junk bonds, 543

K

Kahneman-Tversky prospect theory, 1940, 1942–1945, 1948, 1949, 1957
 Kahneman-Tversky value function, 1942, 1943
 Kalman filter approach, 1601
 Kaplan-Meier estimator, 1922
 Keiretsu, 793
 Keiretsu affiliation effect, 813
 Keiretsu main bank relationships
 data, 797
 firms' equity, 797
 innovation, 793
 model specification, 797, 798
 risk tolerance, 815
 2SLS regression results, 806, 807, 809
 2SLS with instrumental variable, 804, 806
 summary statistics, 798, 800, 801
 univariate tests, 802, 804
 “Kick-in-the-pants” hypothesis, 886
 Kish, R.J., 1398
 KMV Distance-to-Default method, 1313
 Koh, W.T.H., 1398
 Kolmogorov backward equation, 1171
 Korobov filter, 915
 KPMG survey, 2484
 Kreps–Porteus preferences, 562

L

Labor market condition, 2141, 2142, 2144, 2164, 2165, 2167, 2169
 Lagrange multiplier (LM), 1818
 method, 1989
 Lambrecht and Myer's method, 1606
 Large vs. small corporate board, 1197
 Latent Dirichlet allocation (LDA), 2102
 Latent variables, 634
 Latin American debt crisis, 974, 2127
 Lattice method, 911, 914–916, 918, 922, 932
 LCP-GARCH, 2196
 Learning, 1030
 Lease Financing, 2020, 2022, 2024
 Lease transactions
 accounting and financial reporting requirements, 2020
 accounting and reporting requirements, 2032
 accounting guidance, 2038
 arrangements, 2036, 2037
 asset's economic life, 2020
 capital expenditure decisions, 2020
 classification, 2027–2029
 discount rate, 2031
 economic motivations, 2020
 economic theory, 2020
 factors, 2022
 fair value, 2031
 finance literature, 2020
 financial reporting incentives, 2024, 2025
 financial statements, 2025, 2026
 financing, 2023
 IFRS, 2037
 lease term, 2029
 lessee measures, 2032–2034
 lessor, 2034–2036
 nature, 2020, 2021
 non-reporting incentives, 2023, 2024
 payments and residual values, 2030
 U.S. GAAP, 2037
 usage, 2021, 2022
 Lease-versus-buy, 2038
 Least squares (LS), 1801–1807, 1826
 Le Chatelier Principle
 of Markowitz model, 1150–1152
 policy implications of, 1152
 Lee, C.-F., 1398
 Lee and Primeaux's method, 1605
 permanent earnings, 1618
 results from pooled regression, 1613
 statistical information of current and permanent EPS, 1612

- Leland/Leland and Toft (LT) model, 1310
Lender, 2264
Lender of last resort (LoLR), 2257
Less developed countries (LDC) loan, 1179
Less secured jobs, 2143
Leverage, 798, 1642–1647, 1651
Leveraged buyouts (LBOs), 2460
 CLOs, 2473
 debt, 2473
 execution, 2473
 financial techniques, 2472
 investor, 2473
 KKR acquisition, 2472
 LP, 2473
 prevalence, 2479
 risk exposure, 2473
 target firm control, 2473
Liabilities, 2493
Life insurance companies, 698
Life-threatening events, 2140
Likelihood ratio (LR), 1818
Likelihood ratio test, 2577
Limited-information estimation (LML-CCF), 1082
Limited liability, 964
Limited liability company (LLC), 1260
Limited partnership (LP), 1260, 2473
Limit order, 998, 1012, 1103
Limit Order Display Rule, 724, 725
Limits to arbitrage, 1317, 1319, 1320, 1329, 1337
Limits to investor arbitrage, 1319–1320
Linear function, 635, 2636
Linear panel data models, 1802–1806
Linear penalization, 1398, 1399, 1402, 1404, 1407
Linear price-change relationship, 686
Linear probability model (LP), 783
Linear programming model, 1554, 1555
Linear regression, 630
Linear relationships, 1686
Linear simultaneous equation system, 2383
Lintner, J., 1398
Lintner model, 1606, 1607
Liow, K.H., 1398
Lipton, A.F., 1398
Liquidating value, 2493
Liquidation, 775–777
Liquidity, 965, 1002–1004, 1103, 1534, 1536, 1537
 management, 2599, 2621, 2622
 premiums, 607
 ratio, 1002, 1003
 regulation, 2252, 2257
 valuation model, 1496
 risk and return, 1002
Liquidity coverage ratio (LCR), 1244, 1245, 2254
Liquidity mismatch index (LMI), 2260
LISREL methods, 2383–2385
Listed stocks, 1097–1098
Loan commitment contracts, 704
Loan contract terms
 definition, 703
 lending syndicate, characteristics of, 705–707
 measures of risk, 708–709
 simultaneity, 707–708
Loan facilities, 703
Loan market, 1124, 1129–1131
Loan-repayment probability function, 1126–1127
Loans and leases per total asset (L/A), 1995
Local mean–variance efficiency, 749
Local parametric approach (LPA), 2178, 2180, 2192, 2196
Local volatility model, 1904–1907
Logarithmic forms, 1688
Log consumption/wealth ratio, 559
Logistic smooth transition (LSTR), 1689
Logistic transformation, 1689
Logit and probit models, 783
Logit model (LM), 783
Logit regression model, 2163
Log-likelihood function, 1424
Log Likelihood Ratio Statistics, 1695
Log-linear budget constraint, 558–559
Lognormal distribution, 1585, 1595
Lognormal model, 822, 1913, 1915–1916
Lognormal process, 822
Lognormal versus normal movements, 821–822
LOGTA, 798
Log utility, 1949
London Interbank Offered Rate (LIBOR), 574
Longer-maturity loans, 704
Longer-term assets, 513
Long short-term memory (LSTM), 2180, 2186, 2188, 2190, 2196, 2199–2202, 2209
Longstaff and Schwartz (LS) model, 1302
Long-term bank-firm relationship, 793
Long Term Capital Management (LTCM) Ltd, 1267, 1984
Long-term company survival, 1215
Long-term debt, 980
Long-term operating accruals (LOA), 1857
Long-term payout ratio, 1607, 1611–1612

- Long-term trends, 1050, 1051
- Loss aversion, 1946, 1952
- Loss aversion function, 1944
- Low-discrepancy (LD) sequences, 1884–1886
- Lucent New Ventures Group (LNVG), 903
- Luehrman's option space, 900

- M**
- Macaulay duration, 689, 691, 693
 - pre-funded coupon bonds, 544–546
 - zero-coupon bonds, 544–546
- Machine learning, 1872, 1873, 1881, 1889, 1897, 2095–2097, 2178–2180
 - algorithms, 2098–2101
 - applications
 - in accounting research, 2101–2106
 - in auditing research, 2106–2108
 - vs. statistical modeling, 2097–2098
 - tools, 2108–2109
 - trader, 2185–2192
- MacLaurin expansion, 2421
- Macroeconomic environments, 2599, 2619
- Macroeconomic equilibrium, 1125, 1129, 1130
- Macroeconomic factors, cross border M&A
 - current economic conditions in the home country, 1060
 - diversification, 1060
 - exchange rates, 1059, 1060
 - favorable acquisition factors, 1059
 - government restrictions and regulations, 1061, 1062
 - inefficient management, 1061
 - monopolistic power, 1061
 - technological and human resources, 1060
 - unavailability of information, 1061
 - unfavorable acquisition factors, 1060
- Macrohedging, 682, 686, 696
- Make home affordable (MHA) initiatives, 1387
- Managed funds, 668, 674
- Management board, 1191
- Management flexibility, 896, 898, 900
- Management inefficiency, 1064
- Management teams, 2494
- Managerial effort, 1584, 1589
- Managerial extra-effort, 1584
- Managerial incentive scheme, 1587
- Managerial issues, 2120
- Managerial labor market, 2142
- Managerial overconfidence, 1583, 1594
 - compensation contracts and managerial effort, 1584
 - simulation procedures, 1587–1589
- Managers-shareholders' conflict, 1215
 - cash flows, 1215
 - consequences, 1216
 - Horizon problem, 1216
 - risk problem, 1215
- Managing Partner of Long-Term Capital Management, 1252
- Margin, 1704
- Marginal tax rate, 2048, 2049
- Marginal tracking error, 2289
- Margin floor, 1451
- Market
 - approach, 2492
 - concentration, 2482
 - control mechanism, 1211
 - frictions, 952–954, 959
 - hedging risk, 556, 570, 571
 - liquidity supplied by limit order traders, 1103
 - order, 998, 1012
 - portfolio, 622
 - power hypothesis, 879, 884
 - price movements, amplification of, 1455
 - price of risk, 635
 - quality, 1096, 1097, 1101–1104, 1107
 - reaction, 882, 885, 890
 - risk, 686, 2252
 - structure, 1003, 1012, 1013, 1015
 - timing, 640, 643, 644, 647, 648
 - type, 2123
- Marketable securities, 652, 654–657, 660–661
- Market-based governance mechanisms, 732
- Market efficiency, 640, 850
 - definition, 937
 - efficient market model, 938
 - joint hypothesis problem, 939
 - semi-strong-form tests, 941
 - strong-form tests, 941
 - weak-form tests, 940–941
- Market interest rate forecast (MIRF), 1912
- Market makers
 - adverse selection, 1009
 - bid-ask spread, 1009
 - competitive environment of, 1009
 - liquidity, 1008
 - roles, 1008
 - stock exchange, 1008
- Market Value of Equity (MVE), 2492
- Market Value of Invested Capital (MVIC), 2492
- Markit Data Explorers (DXL) equity lending data, 2326
- Markov process, 1305
- Markowitz model, 1150–1152

- Markowitz stochastic dominance (MSD), 862–866
- Matching, 1781, 1782, 1784–1797
principle, 664–665
- Material adverse change, 2479
- Material adverse effect, 2479
- Mathematical method, EIV problem, 2380–2382
- Matlab, 918
- Matrix-weighted averages, 1812
- Maturity, 703–708
- Maturity-bond strategy, 693
- Maximal overlap discrete wavelet transform (MODWT), 2581, 2582
- Maximum likelihood (ML), 1807, 1815, 1818, 1827
- Maximum likelihood estimator (LME), 782
- Maximum likelihood method, EIV problem, 2382–2383
- Maximum tax liability, 1538
- Maximum transfer price, 1556
- Mazumder, M.I., 1398
- McKenzie, M., 1398
- MCM Horse Races, 2294–2296
- Mean reversion, 824, 826
- Mean–variance approach, 744
- Mean–variance efficiency, 621, 744
portfolios, risk and return of, 2284
- Mean–variance frontier, 745, 746
- Mean-variance modeling, 755
- Mean-variance (M-V) rule, 869, 870
- Mean variance portfolio allocation
APT, 749
CAPM, 747–748
estimation errors, 750–751
ICAPM, 749, 750
portfolio choice, 750–751
portfolio selection, 744
- Mean-Variance Tracking-at-Risk (MVTaR) simulations, 2291
- Measurement error, 2370, 2372
in investment equation, 2403–2405
- Measurement errors, regression coefficients, 1621–1625
- Mehra-Prescott equity risk premium, 1945
- Membership function, 2675
- Members' voluntary liquidation, 777
- Memorandum of understanding (MOU), 2468
- Mental accounts, 1945
- Merger, 878, 2460
- Mergers and acquisitions (M&As)
accounting, 887–888, 2493, 2494
agency theory, 881–883
bankruptcy avoidance hypothesis, 886–887
business implications, 2458
corporate events, 2458
definition of, 878–879
different deal structures, 2459
diversification hypothesis, 885
efficiency theories, 879–881
expansionary investment decisions, 2458
experience, 2458
external investment, 2458
features, 2460
financing, 2469, 2470
free cash flow hypothesis, 883–884
goals, 2462
information hypothesis, 886
legal and regulatory framework, 2482, 2483
market, 2458
market power hypothesis, 884
market reaction, 890
motivational factors, 2469
performance, 2495
risk and success, 2494, 2495
stages, 2464
strategic objectives, 2465
strategic rationale, 2464, 2465
tax effects, 887–888
takeover financing and payment, 888–890
transactional forms, 2468
transactions, 2459, 2460, 2470
transaction types, 2460
US company, 2469
valuation, 2485, 2486
- Merton's model, 556, 1360, 1361
- Message Passing Interface (MPI), 1888
- Meta-classifier, 2197
- Mexican CPI inflation, 976
- Mexican moratorium, 1180
- Mexican peso crisis
current account deficit, 976, 978
foreign exchange reserves, 978
inflation comparison, 978
Mexico's balance of payments, 975
nominal exchange rate, 978
Peso/USD exchange rate, 977, 978
purchasing power parity (PPP) based exchange rate, 978
refinancing package, 980
stock index, 979
- Mexican stock index, 979
- M-GARCH model, 1413, 1414
- Microeconomic factors, cross border M&A
composition, 1062
foreign firms to the U.S. market, 1062
international, 1062

- Microeconomic factors, cross border M&A
 (*cont.*)
 synergy hypothesis, 1064–1066
 U.S. companies, 1062
 undervaluation, 1062–1064
 value maximization of the firm, 1066–1068
- Microsoft, 903
- Microstructure/micro-finance approach to
 exchange rates
 currency crisis, 948
 definition, 945, 946
 empirical failure of traditional
 approaches, 946
 heterogeneous information, 947
 information, 948
 information role of order flow, 948–949
 private information, 947
 stock market crash, 948
 trading process of exchange rates, 948
- Miles-Ezzell formula, 1642, 1645–1647, 1649
- Miller and Modigliani theorem, 2410,
 2413, 2417
- MIMIC methods, 2383–2385
- Minimum discrepancy (MD), 674
- Minimum transfer price, 1555, 1556
- Minimum-variance efficient, 621
- Minimum-variance (MV) hedge ratio, 1707,
 1708, 1717
- Minimum-variance portfolio, 621, 632,
 634, 745
- Ministry of Finance (MOF), 1037
- Misclassification cost, 2187, 2188
 model, 786
- Mixed-jump processes, 1075
- Modern portfolio theory, 743, 2278, 2279
- Modified immunized portfolio, 688
- Modigliani, F., 1400
- Modigliani, L., 1400
- Modigliani and Modigliani measurement,
 988, 989
- Modigliani and Modigliani's index, 1398
- Modigliani-Miller formula, 1644, 1645, 1649,
 1650
- Modigliani-Miller model (MM), 880
- Modigliani-Miller theorem, 2023
- Modigliani-Modigliani measure, 988
- Moment restrictions, 1174
- Momentum crashes, 1119
- Momentum effect, 1110
- Momentum profits, 1117
 country, 1118
 firm, 1117, 1118
 market states, 1119, 1120
- Momentum strategy, 1118–1120
- Monetary Authority of Singapore (MAS), 1139,
 1145, 1146
- Monetary base, 1131
- Monetary neutrality, 1133, 1134
- Monetary policy(ies), 1044, 1047, 1052, 1124,
 1125, 1130–1135
 operations, 2260
- Money center banks, 1984
- Money returns (MR), 1860
- Monitoring, 968
- Monopolies and Mergers Commission
 (MMC), 884
- Monopolistic power, 1061
- Monotonicity axiom, 852
- Monte Carlo method, 911–913, 918, 923, 932
- Monte Carlo simulation (MCS), 1873, 1875,
 1882, 1891, 1892, 1894, 1896, 2488
 Monte Carlo and Quasi-Monte Carlo
 methods, 1884–1885
 option pricing, 1885–1886
 VaR, Monte-Carlo bootstrap for, 1886–1887
- Moody's-KMV model, 1297
- Moody's-KMV (MKMV) structural credit risk
 model, 1297–1299
- Moral hazard, 512, 514, 515, 524, 881
- Morgan Stanley Capital International World
 Index, 1276
- Mortgage
 amortization schedule, 2265
 cash flows, 2265–2268
 definition, 2263
 factors, 2268
 functions, 2267, 2268, 2274
 length, 2264
 payments, 2267
 principal, 2264
 rate, 2264, 2265, 2268
 sensitivity, 2268, 2274
 WAL, 2269, 2270
- Mortgage-backed securities (MBS), 836
 characteristics, 1158
 default-free models, 1160
 endogenous termination, 1163–1164
 exogenous and endogenous
 prepayment, 1160
 exogenous prepayment, 1163
 expected termination level of pool, 1172
 expected termination probability, 1170
 generalized method of moments, 1172, 1173
 heterogeneous mortgages, 1165
 housing price process, 1160
 issue, 1159

- moment restrictions, 1174
 - mortgagors' heterogeneity, 1159
 - one-factor model, 1162
 - option pricing theory, 1159, 1161, 1163, 1175
 - pooled mortgage liabilities, 1169
 - pricing, 845–846
 - security price, 1170
 - story, 844
 - termination decision of a single mortgagor, 1166–1169
 - termination probability of mortgage, 1171
 - transaction costs, 1165
 - two-factor model, 1161
 - valuation model, 1161
 - Mortgage insurance premium (MIP), 1377
 - Mortgagor, 2264
 - liabilities, 1163
 - Multi-beta asset pricing models, 619–621
 - Multi-beta conditional asset pricing models, 635–636
 - Multicollinearity, 1809
 - Multidimensional integration, 911
 - Multi-factor models, 692
 - Multifactor pricing models, 749
 - Multilayer perceptron (MLP), 2180, 2186, 2188, 2189, 2196, 2199–2202, 2209
 - Multi-movement model, 1501
 - Multinational companies (MNCs), 2484
 - Multiple-beta models, 635, 636
 - Multiple criteria linear programming (MCLP)
 - classification, 1768–1771
 - data collection and research design, 1771–1776
 - data mining approach, 1767
 - Multiple-factor model (MFM), *see* Barra risk model
 - Multistage real compound call option, 907–909
 - Multivariate discriminant analysis (MDA), 779
 - Multivariate GARCH (MGARCH) models, 2558, 2561, 2585
 - Multivariate normal integral, 909, 911, 913, 914, 916, 918, 932
 - Drezner's method, 913
 - lattice method, 914–916
 - Monte Carlo method, 912, 913
 - numerical methods, 918
 - Mundell-Fleming model, 1135
 - Mutual fund industry, 1283
 - Mutual funds, 640, 642, 644, 646, 647
 - M-V efficiency analysis, 869, 871, 874
- N**
- NA performance bounds, 672, 673
 - Narrow framing, 1946
 - NASDAQ markets, 720, 758
 - Nash bargaining wage, 1141
 - Nash equilibrium, 1145, 1146
 - Nash game, 1145
 - National Association of Securities Dealers (NASD), 723, 1099
 - National Association of Securities Dealers Automated Quotation (NASDAQ) system, 723–726, 1094, 1095, 1098–1102, 1107
 - National Center for High-Performance Computing (NCHC), 1890
 - National Institute of Standards and Technology (NIST), 1874
 - National Market System (NMS), 1098
 - National Quotation Bureau, 1099
 - National Wages Council (NWC), 1138, 1139, 1143–1147
 - Natural disasters, 1780, 1782
 - Nature's methods, 882
 - Negative convexity, 686
 - NegEarn_{j,t}*, 2509
 - Negotiated transfer pricing (NTP), 1553
 - Negotiated wage, 1140
 - Net asset approach, 2486
 - Net interest cash flows (INTt), 1854
 - Net Liability of Firm in Pension Plan (NLFPP), 2044
 - Net present value (NPV), 1399, 1404, 1405, 1407, 2464, 2657, 2668, 2674
 - analysis, 1066, 1067
 - framework, 896–899, 901
 - Net stable funding ratio (NSFR), 1244, 2254
 - Netstock, 756
 - Net working capital, 653
 - Networks, 1094
 - for listed stocks, 1097–1098
 - OTC, 1099–1101
 - priority rules, 1101
 - Neural network(s), 1031, 2101, 2180, 2186–2188, 2192, 2196, 2197, 2199, 2209
 - News-based exchange rate volatility hypothesis, 590, 591
 - Newton–Raphson method, 916, 1512
 - New York Stock Exchange (NYSE), 720, 724, 726, 1094, 1095, 1097, 1098, 1101–1103
 - market makers, 1014
 - stock, 940

- Nikkei stock index, 2178
 No-arbitrage approach, 953, 959
 Nodes, 1094, 1101, 1107
 No-extra-effort and certainty equivalent (NECE), 1592
 Nokia Group, 903
 Nolan maximum likelihood technique, 1991
 Nominal exchange rate, 1140
 Nominal wage, 1140
 Non-audit services, 1199
 Nonbankrupt industrial companies, 782
 Non-cash compensation, 1212
 Nonclassical maximum likelihood estimator (NMLE), 785
 Non-disclosure agreement (NDA), 2468
 Nondistressed and distressed bidders, 890
 Nonlinear models, 1686, 1687, 1691, 1697
 Nonlinear regression, 1697
 Non-Nash optimal appreciation rate, 1146, 1147
 Nonnegotiated takeover bids, 734, 735
 Nonseparable-consumption models, 618
 Non-South Korean exposed bank group, 1182
 Non-surviving entity, 2468
 Non-traded asset, 901, 925
 Nordbank, 1984
 Norddeutsche Landesbank, 1984
 Normal distribution, 1986
 of returns, 1399, 1403
 Normal market, 1985
 Normal model, 822, 1915
 Normal vibrations, 1075
 North American Free Trade Agreement, 974
 “No-shop” provision, 2479
 No short sales, 953, 955, 956, 958
 No tax shields, 2053
 N test assets, 678
 Null hypothesis, 748
 Numerical methods
 partial differential equations, 1575–1577
 tree approach, 1577
- O**
- Obfuscation, 1218
 Occupational pension system, 2042
 Office of Economic and Corporate Development (OECD), 1188
 Offset coefficient, 1132, 1134
 One-day out-of-sample volatility forecast loss function, 2587
 “One-dimensional” thinking, 899
 One-factor model, 1162, 1989
 One-month forward exchange rates, 587
 One-period binomial model, 956
 One-step statutory merger, 2469
 Online trading
 Campbell–Shiller model, 757
 Cornell model, 757
 day-trading centers, 753
 Fed model, 757
 full-service brokers, 754
 Glassman–Hassett model, 757
 individual investors, 754
 issues, 754–756
 Opaque, 1218
 Open limit order book auction, 994
 Open market operation, 1130, 1131, 1133, 1135
 “Open outcry” method, 1095
 Operating accruals (OA), 1851, 1855
 Operating leases, 2021, 2022, 2025–2028, 2033, 2035–2038
 Operating risk, 2070, 2504, 2507
OperCycle_{j,t}, 2509
 Optimal, 1544, 1547, 1549, 1550, 1552, 1554, 1560
 corporate pension policy, 2051
 cut-off point, 785–787
 pension policy, 2056
 weighting matrix, 676
 Optimal hedge ratio (OHR), 1411, 1413
 Optimal payout ratio, 2420
 empirical evidence, 2430–2431
 and growth rate, 2420–2426
 investment, financing and dividend policy, 2442–2444
 moving estimates process for structural change model, 2440, 2451
 multivariate analysis, 2436–2440
 no change in risk, 2429–2430
 and risks, 2426–2430
 statistics for sample firms, 2431–2432
 systematic risk, 2428
 total risk, 2426–2427
 total risk and systematic risk, 2428–2429
 univariate analysis, 2435–2436
 Optimum dividend policy, 2418–2420
 Optimum mean-variance hedge ratio, 1708
 Option-adjusted spread (OAS), 1574
 Option-based portfolio insurance (OBPI), 1440–1441, 1449, 1459–1461, 1463
 Option pricing, 1873, 1875, 1882–1886, 1891, 1896, 1897
 model, 1512–1514, 1526
 theory, 1159, 1161, 1163, 1175
 Order batching, 995–998

- Order driven facility, 992
 - Order-driven market, 1014
 - Order flow, 948–949
 - Order-handling rules (OHRs), 1014
 - Order-processing costs, 1009
 - Ordinary least square (OLS)
 - conditional covariance/variance, 1712
 - conventional regression model, 1716
 - error correction estimates, 1717
 - heteroscedastic errors, 1716
 - maximum likelihood estimator, 1712
 - random walk, 1712
 - regression, 1711
 - R language, 1723
 - summary statistics, 1715
 - technique, 1711
 - unit roots and residual cointegration tests, 1717
 - vector of variables, 1712
 - Organizational inefficiency, 883
 - Organization for Economic Cooperation and Development (OECD), 1058
 - Ornstein-Uhlenbeck process, 1742
 - O-scores, 794, 809
 - Other equities per total asset (O/A), 1995
 - Other operating accruals (OOA), 1858
 - Out-of-sample hedging performance, 1427–1430
 - Outside directors, 1213
 - nomination, 1196
 - Outsiders, 1192
 - Outside shareholders, 1194
 - Overconfidence, 1582, 1583, 1585, 1587–1592, 1594, 1595
 - level, 1588, 1590, 1591
 - measure, 1592, 1593
 - model, 1117–1119
 - Overlapping-generations (OLG), 1355
 - Overreaction, 1114–1116, 1119
 - Over-the-counter (OTC), 723, 1099–1101
 - Ownership structure, 968
- P**
- Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) grid, 1873, 1892–1893
 - Pac-Man tactics, 2477
 - Panel data, 2214
 - Parametric method, 1987
 - Parametric models, 1933, 1934
 - Paretian distribution, 1363, 1985, 1986, 1990, 1992
 - Paris Bourse, 1095
 - Partial adjustment coefficient, 1611–1612, 1615
 - Partial-adjustment model, 1600
 - Partial differential equation (PDE), 901, 902, 908
 - Partial structural change model, 2216
 - Past losers, 1109, 1114, 1118
 - Past winners, 1109, 1114–1116, 1119
 - Pay concession, 2141
 - Payoffs, 953
 - Payout policy, 1687, 2410, 2411, 2413, 2430, 2431
 - Payout ratio, 1625–1639, 2357
 - Payroll taxes, 2048
 - Pecking order theory, 889
 - Penalized Internal Rate of Return (PIRR), 1398–1400, 1402–1404, 1407, 1408
 - Penalized Internal Rate of Return for beta (PIRR for beta), 1401, 1402, 1404
 - Penalized Present Value (PPV), 1399, 1404–1407
 - Penalized Present Value for beta (PPV for beta), 1406
 - Pension accounting, 2045, 2047
 - fair-value, 2044
 - inside debt and capital structure, 2042
 - Pension actuarial assumptions, 2045, 2046
 - Pension beneficiaries, 2051
 - Pension Benefit Guaranty Corporation (PBGC), 2043, 2044
 - Pension collateralized debt, 2049–2051, 2054
 - Pension contract, 2048
 - Pension contracting process, 2050
 - Pension demand, 2049
 - Pension funding policy, 2054
 - Pension funds, 646–648, 698, 699
 - Pension information, 2046, 2047
 - Pension plans
 - accounting under FAS 158 and IAS 19, 2044
 - and capital structure decisions, 2047–2049
 - DB plans (*see* Defined benefit (DB) plans)
 - DC plans (*see* Defined contribution (DC) plans)
 - NLFPP, 2044
 - PBGC, 2044
 - Pension policy, 2056
 - Pension-related liabilities, 2045
 - P/E ratio, 756, 2357, 2364
 - Percentage sales method, 2344–2345
 - Performance-based incentive plan, 2147
 - Performance evaluation techniques, 987
 - Performance indexes, 1398

- Performance measures, 762, 765, 768
- Permanent dividends, 1603
- Permanent earnings, 1601, 1607, 1611, 1612, 1618, 1620
- Permanent EPS, 1602, 1605, 1610, 1612
- Permanent income, 1621
- Permanent-income hypothesis, 1602
- Peso/USD exchange rate, 977
- Phantom income, 605
- Phillips-Perron (PP) test, 2563
- PicsouGrid, 1875
- Piecewise regression, 1694, 1695
- Pink Sheets, 1099
- Poison pill, 735, 1212, 2476
- Poisson–Gauss model, 1080
- Poisson process, 1074–1076, 1080, 1082
- Policy-game player, 1144
- Pooled mortgage liabilities, 1169
- Pooling and purchase methods, 888
- Portfolio(s)
 - choice, 750–751
 - diversification, 858–860
 - duration, 687
 - formulation strategies, 754
 - lambda, 2291
 - performance, 688
 - risks and returns of, 2277–2278
 - theory, 860
 - tracking error, 2288
 - weights, 642, 644, 645
- Portfolio insurance
 - definition, 1438
 - empirical comparison, 1458, 1463
 - expected shortfall, 1447
 - financial market stability, 1455
 - history, 1449
 - market developments, 1447
 - theory, 1439, 1440, 1447
- Portfolio performance evaluation, 984
 - benchmark index, 984
 - broader market index, 984
 - conventional method, 984
 - growth style, 984
 - investment portfolio, 984
 - management, 984
 - style-comparison, 984
 - style-oriented funds, 985
 - value-oriented portfolio, 984
- Portfolio performance measures
 - Friedman test, 2630
 - ranking of random portfolios, 2628–2629
- Portfolio risk management
 - application to Value-at-Risk, 2576–2581
 - conditional variance and volatility analysis, 2568–2574
 - data for empirical analysis, 2562
 - descriptive statistics, 2562, 2564
 - dynamic conditional correlational analysis, symmetric and unsymmetric, 2569–2574
 - econometric model specification, 2564–2567
 - forecast performance evaluation, 2574–2576
 - unit root tests, 2564
 - univariate GARCH analysis, symmetric and asymmetric, 2568–2570
 - wavelet-based approach, 2580–2590
- Portfolio-specific tracking error, 2289
- Portfolio Value-at-Risk (VaR), 2290
- Positive convexity, 686
- Post-closing integration, 2481
- Post-consolidation keiretsu banks, 794, 815
- Post-merger integration (PMI), 2458, 2494, 2496
- Pre-approval requirement, 1199
- Predictability, 629, 630, 632–634, 636
- Preempt competition, 890
- Pre-funded coupon bonds
 - asset-based coupon payments, 550
 - bond interests, 543
 - cost of funding, 543
 - coupon payments, 544
 - default risk, 544
 - escrow return, 553
 - financial engineering, 542
 - financial market, 546
 - first issued in 1994, 541
 - interest subsidy, 550–552
 - loss of value, 550, 551
 - market conditions, 544
 - market price, 544
 - Mccauley duration, 544–546
 - net duration, 547, 549–553
 - 5 or 10 year maturity, 546
 - pre-funded issue, 542
 - reinvestment risk, 544
 - risk characteristics, 543
 - total face value, 547, 548, 553
 - value loss, 544–546, 552
 - 5-year, 547
- Premium burnout effect, 1159
- Prepayment, 845
 - benchmark, 2271
 - CPR, 2271
 - PSA, 2271–2273
 - SMM, 2270

- Prepayment speed assumptions (PSA), 2271, 2274
- Preservation, 2494
- Press, 968
- Pre-tax profits, 1839
- Price and time priority, 997
- Price determination, 995–998
- Price discovery, 609, 610, 995, 1000, 1008, 1013, 1015
- Price elasticity, 684, 689
- Price formation, 937, 938
- Price improvement, 993, 997
- Price scan auction, 994
- Price stabilization, 1008
- Primary restructuring approaches, 2471
- Primitive assets, 676
- Principal–agent conflicts of interest, 964–966, 968, 970
- Principal components analysis (PCA) models, 2277
- Principal protected note (PPN), 579
- Priority rules, 1094, 1096–1098, 1100, 1101, 1104, 1107
- Private banking system, 1235
- Private benefits of control, 966
- Private equity (PE), 2458
- Private/proprietary reverse mortgages, 1374
- Private universities, 1967
- ProActive, 1875
- Probability class, 2185, 2190, 2192, 2195
- Probability density function, 1569
- Probability distortion, 850–852, 869, 872, 873
- Probit model (PM), 783
- Process Driven Trading (PDT), 1268
- ProMos, 897, 898, 924, 928, 930
- case study, 929
- dividend-like yield, 928
- DRAM technology, 925, 928
- exercise prices, 928
- model, 925
- valuation, 931, 932
- volatility, 930
- Propensity score matching (PSM), 812, 813, 1784, 1786, 1789, 1792–1796
- Prospective reference theory, 854
- Prospect stochastic dominance (PSD), 862
- Prospect theory (PT), 852, 854, 855, 869
- Proxy advisory firms, 2476
- Proxy fight, 2468
- Proxy mechanism, 2468
- Proxy statement, 733, 734
- Prudential ICICI Mutual funds, 1279
- Psychological bias, 1582
- Public company merger agreements, 2479
- PUBLICDEBT, 798, 813
- Public debt on innovation, 795
- Public debt ratio, 793, 798, 802, 804, 806, 808, 812, 813, 815
- Public Securities Association (PSA), 845
- Public universities, 1967
- Purchase price allocation (PPA), 2494
- Purchasing power parity (PPP), 556
- based exchange rate, 978
- Pure discount bonds, 542
- Pure structural change model, 2216
- “Pure structural” model, 1733
- Pure style factor portfolios, 2287
- Put option, 1882
- PV() function, 2268
- Pyout ratios, 2436
- Q**
- Quadratic programming, 1150
- Qualified utility functions, 1949–1950
- Quantitative easing (QE), 1997
- Quasi-maximum likelihood estimation (QMLE), 2194
- Quasi-Monte Carlo (QMC) methods, 1884
- Quote-driven market, 1014
- R**
- Race-to-the-bottom phenomenon, 731
- Race-to-the-bottom theory, 734
- R&D expenditure, 792, 815
- ratio, 801
- Radical equations, 1950
- Radical functions, 1950, 1960
- Radical utility function, 1950–1951, 1958
- R&D investment, 1689–1691, 1694
- with one switching point, 1692
- Rainbow options, 1887
- Raising capital (RC), 1995
- Random coefficient model, 1183
- Random effects models, 1810–1813
- Randomness, 2660
- Random priority, 1095
- Random stock, 1104
- Random variables, 1801
- Random walk model, 938–940
- Rank correlation, 762, 763, 767, 768
- Rank dependent expected utility (RDEU), 850, 853, 854, 872, 874
- Ranking(s), 762, 764, 768, 1400, 1402
- Rare-event equilibrium model, 1985

- Ratchet floor, 1450
- Rate distributions, 1907, 1911
 - mean and median of, 1911–1913
- Rational beliefs equilibrium (RBE), 1361
- Rational expectations, 586, 630, 634
- RBNB DataTurbine, 1879, 1895
- Real Business-Cycle Theory, 1348
- Real call option, 908, 918, 920, 922–924, 932
- Real earnings management, 2066, 2067, 2069, 2071–2074, 2077–2079, 2084–2087
- Real Estate, 836, 838, 847
- Real Estate Investment Trusts (REITs), 836–839, 843–845, 847, 848
- Real interest rate, 586, 592–598, 607
- Realization utility, 1946
- Real options, 2659
 - approach, 2658
 - dividend-like yield, 902
 - expanded NPV, 898
 - investment project, 899
 - Luehrman's option space, 900
 - nontraded assets, 901–902
- Real option value (ROV), 2659, 2661, 2662
- Recession, 1146
- Recessionary periods, 695
- Recovery price, 575
- Recursive preference, 563
- Reference asset, 575
- Reference entity, 575
- Refinancing transaction costs, 1159–1161, 1164, 1165, 1169, 1172, 1175
- Regime switching, 1500
- Regions financial
 - capital ratios, 2008, 2014
 - estimation, 1996
 - security assets, 2003
- Reg SHO data-NYSE Short Sales (TAQ), 2325
- Regulation, 515
- Regulatory capital, 2127–2130
- Reincorporation
 - agency costs, 730
 - competition, 730–731
 - director liability, 736–738
 - promoting shareholder rights, 735
 - substantial transition, 733
 - tax savings, 738
- Relationship-based bank financing, 792
- Relative risk aversion (RRA), 556, 557, 562, 566, 568, 571, 1943, 1954, 1956
- Renminbi Special Shares, 714
- Representative heuristic, 1115
- Required capital ratio (RCR), 1989
- Required stable funding (RSF), 2254
- Reserves, 889
- Residual income (RI), 1546
- Resiliency, 1002, 1003
- Responsibility centers (RC)
 - cost centers, 1544
 - decision-making, 1543
 - investment centers, 1545
 - profit centers, 1544, 1545
 - residual income (RI), 1546
 - Return on Investment (ROI), 1545, 1546
 - revenue centers, 1544
- Retained earnings, 1600, 1603
- Return distribution, 1364
- Return on assets (ROA), 2601
- Return on Investment (ROI), 1545, 1546
- Return predictability, 2320–2321, 2332
- Returns-based measures, 644
- Reverse mortgage, 1374
- Reverse termination fee, 2479
- Reverse triangular merger, 2469
- Revlon case (1985), 2481
- Risk(s), 1018–1024
 - assessment, 1219
 - bearing, 1009
 - disclosures, 1219
 - dynamics, 640, 642
 - measure, 2180, 2181, 2184
 - seeking, 2143
 - structure of interest rates, 1296
 - taking/aversion, 2142
- Risk-adjusted methods, 985
- Sharpe ratio, 985
- Treynor ratio, 986
- Risk-averse CEOs, 2143
- Risk aversion, 556, 557, 561, 562, 566, 568, 571, 617, 856, 861, 862, 870, 1348–1354, 1358, 1360, 1365–1369, 1587, 1945, 2181, 2182, 2634–2636, 2641, 2642, 2644–2647, 2649–2652
- CARA case, 2642–2645
- hypothesis, 1690
- Risk-based portfolio insurance (RBPI), 1443
- Risk-free asset, 623, 746
- Risk-free rate, 923
 - puzzle, 1348, 1359
- Risk-free return, 632
- Riskless-arbitrage argument, 692
- Risk management, 1873, 1876, 1877, 1897
 - measure of (*see* Value-at-Risk (VaR))
- Risk-neutral default probability, 1303
- Risk-neutral evaluation method, 901
- Risk-neutral pricing, 2638

Risk preference, 2141–2144, 2148, 2149
measures, 2148

Risk treatment

advantages of linear penalization,
1402–1404

PPV, 1404–1407

Sharpe's classification, 1399–1401

Treynor's classification, 1401–1402

Risky assets, 1987

Risky cash flow, 2657

Risky debt, 1296, 1299, 1306, 1311

Robustness checks

alternative innovation measure, 813, 814

exogenous shocks, 809, 810, 812

propensity score matching, 812

Robust standard errors, 2171

ROC curve, 2189, 2198

Root mean squared errors (RMSEs), 2549

Rounding, 721

Rutgers Business School, 1250

S

Safety covenants, 1299

Safety-first rule, 1986

Safety net, 514

Sainz, J., 1398

SaleGrowth_{j,t}, 2508

Santibáñez, J., 1404

S&P 500 Index, 757, 1412, 1413, 1415, 1420

Sarbanes–Oxley Act, 1023, 2483

Sarbanes–Oxley Act 2002 (SOX), 1188, 1190,
1196, 1207

Savings and loan holding companies
(SLHCs), 533

Scale spillover effect, in volatility, 2585

Scikit-learn, 2109

Score matching analysis, 812

Z-score model, 779

Sealed bid auctions, 994

Secant method, 916, 918

Second-order Taylor expansion, 559

SEC Order Handling Rules, 724

Securities and Exchange Commission (SEC),
968, 1670, 2021, 2482

Securities and Exchange Committee (SEC), 726

Securities Exchange Act, 2483

Securities per total asset (S/A), 1995

Securities traders, 968

Securitized Asian Corporate Bonds, 1041, 1042

Security assets, 1994

Security lending process, 2316

Security market line (SML), 2279

Security price, 1170

Security returns, 939, 940

Security selection, 640

Security transaction tax (STT)

earliest proponents of, 1532

liquidity, 1534, 1536, 1537

volatility, 1532–1536

volume, 1536

Selective targeting policies, 1048

Self-attribution, 1115, 1118

Semi-parametric hazard models, 1925

Semi-strong efficiency, 890

Semi-strong-form tests, 941

Senate Investigating Committee, 1022

Sensitivity analysis, 1594, 2357–2358
dividend-like yield, 920

of optimal payout and growth rate,
2422–2425

of risk-free rate, 923

volatility, 920, 922, 923

Separate Trading of Registered Interest and
Principle of Security Program
(STRIPs), 543

Separation of ownership and control, 964, 966,
967, 970

Separation Theorem, 871, 874

Serial correlation tests, 940

Serial dependence, 633

Severance, 2146

Severance agreement, 2141, 2144, 2146

SFAS-95, 1659–1661, 1671, 1673, 1680–1682

Shanghai Stock Exchange (SHSE), 713, 714

Shanken correction, 1480

Shapely value, 1555

Shareholders, 731–736, 738–740

of corporations, 1324–1329

rights plan, 2475

wealth, 1212

wealth maximization theory, 1068

Share value, 1002

Sharing/pro rata priority, 1095

Sharpe, W.F., 1398, 1399

Sharpe and Information ratios, 2629

Sharpe index, 763, 766

Sharpe–Lintner Capital Asset Pricing model,
748, 1076

Sharpe–Lintner theory, 942

Sharpe ratio, 746, 747, 985, 987, 1399, 1400,
1403, 1404

Shearson Loeb Rhoades (SLR), 881

Shell company, 2473

Shenzhen Stock Exchange (SZSE), 713, 714

- Shock jump intensity residual, 1086
- Shorter-maturity loans, 704
- Shorter-term liabilities, 513
- Short interest, 2314–2317, 2319–2323, 2327–2334
 - information role, 2331
 - short interest, 2324
- Short last lowers (WML), 1119
- Short sales constraints, 953, 956–960
- Short selling, 2313–2318, 2320–2332, 2334
 - consequences, 2323–2324
 - constraints, 2317–2318
 - costs and rebate rates, 2318
 - cross-sectional return predictability, 2320–2321
 - data and empirical evidence, 2325–2331
 - endogeneity problem, 2332–2333
 - institutional structure and activities, 2315–2316
 - interest and corporate events/firm performance, 2321–2323
 - international evidence, 2333
 - market activities, 2316–2317
 - metrics, 2327–2328
 - security lending process, 2316
 - short ban in history, 2324–2325
 - supply, 2320
 - theory, 2319–2320
 - time series return predictability, 2321
- Short-term debt, 976, 978
- Short-term interest rate, 632, 633
- Short-term noise traders, 1533
- Short-term profit, 1215
- Signaling mechanism, 970
- Signal model, 2002
- Silver Parachutes, 1212
- Simplified MIMIC model, 2384
- Simulative analysis, parameter value, 917
- Simultaneity, 707–708
- Simultaneous volatility (SVL) equations, 1414
- Single-beta pricing model, 620
- Single-factor duration measure, 695
- Single-index beta estimation approach, 2280
- Single Month Mortality (SMM), 2270, 2274
- Single mortgagor termination decision, 1166–1169
- Single purpose reverse mortgages, 1374
- “Sitting-on-a-gold-mine” hypothesis, 886
- Sixteenths Minimum Increment Rule, 725
- Size priority, 1095
- 2SLS regression results, 807, 810, 815, 2443
- 2SLS with instrumental variable, 804, 806
- Small Order Execution System (SOES), 723
- Smith vs. Van Gorkom*, 736
- Smooth-pasting condition, 1311
- Social impact bonds (SIB), 1965, 1966
- Social impact investment (SII), 1965–1970, 1972
- Social learning, 1032
- Sorting stocks, 622
- South Korea, 1179, 1181–1183
- South Korean exposed bank group, 1182
- Special Purpose Acquisition Companies (SPACs), 2459
 - attractive merger partners, 2459
 - IPO filing, 2459
 - sponsor, 2459
- Special purpose entity (SPE), 1020–1022
- Special purpose vehicle (SPV), 1040
- Specification analysis, 1601
- Specification error, 2370
- Specification tests, 1801, 1816
- Speculative profits, 940
- Spin Offs Tax Effects, 2471
- Split offs, 2471
- Split ups, 2472
- Spot exchange rate, 586
- Square root function, 1949
- Stable distribution, 1347, 1362–1364, 1368
- Stable model, 1990, 1991
- Stacking, 2196–2199
- Staggered boards, 1212
- Stakeholders welfare measure (SWM), 1592–1594
- Standard cost-based transfer pricing (SCTP), 1553
- Standard deviation, 985, 987
- Standard errors, 806
- Standard exchange rate determination approaches, 946, 947
- Standard normal density function, 1311
- Standard prepayment model, 2271, 2274
- State anti-takeover laws, 2483
- State-based sampling technique, 784
- Statement of cash flows (SCF), 1658
- Statement of Financial Accounting Standards (SFAS), 1658, 2493
- Statement of Financial Accounting Standards 158 (SFAS 158), 2045
- State-time dependent volatilities, 1500
- Statistical methods, 1920
- Statistical modeling, vs. machine learning, 2097
- F*-statistics of moving estimates processes, 2451
- Statman model, 1285
- Statutory mergers version, 2468, 2469

- STDSale*_{*j,t*}, 2509
 Stochastic differential equation, 1306
 Stochastic discount factor (SDF), 614–618,
 620, 623, 624, 630, 632, 635, 636,
 641, 642, 668–676, 678, 1351
 evaluating performance, 668–671
 GMM, 677
 issues, 674–676
 N test assets, 678
 Stochastic dominance (SD) rules, 862, 868, 869
 Stochastic dynamic optimization model,
 2412, 2414
 Stochastic process risk, 682, 690, 691, 693
 Stock, 717
 Stock Exchange of Hong Kong, 1095
 Stock-for-stock offer, 2470
 Stockholders' equity value (SEV), 1586,
 1589–1594
 Stockholders' interest, 966, 967, 969
 Stock market, 1027
 index, 633
 Stock ownership, 714
 Stock prices, 631
 Stock returns (RD), 587, 1074, 1075, 1077,
 1083–1085, 1087, 1119
 Stock risk drivers, 1486, 1488, 1506
 Straight-line floor, 1451
 Strategic alliances, 2473
 Strategic buyers, 2479
 Strategic flexibility, 898, 900, 907, 920, 932
 Strategic goals, 2458, 2459, 2465
 Stress tests
 deficiency, 1985
 externality and signal, 1985
 potential loss, 1993
 principle, 1984
 RCR, 1989
 regulators, 1984
 required capital ratios, 1988, 1989
 risk of assets, 1985
 total assets, 1987, 1988
 VaR models, 1987, 1993
 STRIPS bonds, 820
 STRIPS market, 829
 Strong-form tests, 941
 Structural change/breaks, in panel data
 errors, 2217–2218
 model, 2216–2217
 Structural change/breaks, 2214
 break date estimator, 2219–2220
 objectives, 2214
 testing for, 2218–2219
 in trade credit provision, 2220–2225
 Structural credit risk models
 Black-Scholes-Merton (BSM) model, 1295
 endogenous default model (*see* Endogenous
 default model)
 exogenous default model (*see* Exogenous
 default model)
 Moody's-KMV (MKMV) model, 1297
 Structural models, 1565, 1566, 1570–1573
 Submartingale, 938–940
 Subprime mortgage crisis, 1986
 Summary statistics, 799
 SunTrust
 capital ratios, 2009, 2015
 security assets, 2003
 Supermajority, 1212
 Super martingales, 953, 956
 Supervisory Board, 1191
 Supervisory Capital Assessment Program
 (SCAP), 1984
 Supervisory metrics, 1987
 Supplemental short interest file (compustat),
 2326
 Supply shocks, 1138
 Support Vector Machine (SVM) methods, 1768
 Sustainable growth rate, 2346
 Switching strategy, 2205, 2209
 Symbiosis, 2494
 Syndicated loans, 706, 707
 Syndicate size, 706, 707
 Synergy, 880, 1064–1066
 bidding price, 2487
 higher sales, 2463
 mergers, 2462
 sources, 2464
 Systematic risk, 616, 986, 1399, 1401, 2428
- T**
 Tail index, 1986
 Tail loss optimal, 2190–2192
 Tail risk, 1987
 Tail-risk protection
 auto-regressive models, 2192–2196
 backtest performance, 2200–2206
 dynamic tail risk protection strategy, 2182
 hidden layers architecture, 2189–2190
 investors' preferences, 2181–2182
 model risk evaluation, 2198–2200
 neural networks, 2186–2188
 probability threshold, 2190–2192
 stacking, 2196–2198
 strategy parameters and final decision,
 2190–2192

- Tail-risk protection (*cont.*)
 - tail risk target, 2184–2185
 - target VaR, 2183
 - trader's decision, 2185
 - trader's function mapping, 2186–2188
 - volatility, 2180–2181
- Taiwan Stock Exchange (TSE), 929
- Taiwan system, 1880
- Takeover, 878
- Takeover financing and payment, 888–890
- Tangible common equity ratio (TCER), 1239
- Taper tantrum, 1052
- Targeted stock, 2472
- Target VaR, 2179–2183, 2192, 2198, 2201, 2202
 - strategy, 2181, 2182, 2200, 2203, 2204
- Target zone exchange rate, 976
- Taxable income, 2540–2545, 2547, 2549, 2550
- Tax effects, 879, 887–888
- Tax-free reorganization, 2471
- Taxing exchange-traded instruments, 1537
- Tax payments (Taxt), 1854
- Tax Reform Act of 1986, 1062
- Tax rules, 2541, 2542, 2550, 2551
- Tax shields
 - barrier options, 1642
 - definitions, 1643, 1644
 - discrete-time valuation models, 1642
 - dynamic debt policies, 1650–1655
 - enterprise value, 1642
 - example, 1647
 - financing effects, 1655
 - fixed debt/leverage ratio, 1648–1650
 - formulas, 1643, 1644
 - Miles-Ezzell formula, 1645–1647
 - Modigliani-Miller formula, 1644, 1645
 - notation, 1643, 1644
 - required asset return, 1647, 1648
- Taylor formula, 1516, 1520
- Taylor series expansion, 1956–1957
- Technological innovations, 902
- Tender offer, 2483
- TensorFlow, 2109
- Terminal value (TV), 2488, 2489
- Term structure
 - arbitrage-free models (*see* Arbitrage-free models)
 - equilibrium models, 825–827
 - interest rate correlations, 823
 - of interest rates, 1486, 1501
 - lognormal versus normal movements, 821–822
 - mean reversion, 824
 - movements, 1487, 1501
 - volatilities, 823–824, 1917
- F* tests, 2440
- Thansin Shinawatra, 1037
- The 1997 NASDAQ trading rules
 - Actual Size Rule, 724, 725
 - bid–ask spreads, 726
 - the ECN Rule, 725
 - ESR, 725
 - implementation schedule, 724
 - informed trading costs, 726
 - Limit Order Display Rule, 724, 725
 - market-making costs, 726
 - market reforms, 726
 - NYSE, 726
 - SEC, 726
 - SEC Order Handling Rules, 724
 - Sixteenths Minimum Increment R, 725
 - Sixteenths Minimum Increment Rule, 724
- The Actual Size Rule, 724, 725
- The Merger Process and Decision Making, 2466
- Theory of finance
 - Allias paradox, 851
 - alternative shapes of the utility/value function, 857
 - change of wealth, 855
 - FSD, 852–855
 - integration of cash flows, 855, 856
 - investors' characteristics, 849
 - market efficiency, 850
 - probability distortions, 850–852
 - rational investors, 849
 - risk seeking segment of preferences, 856–858
- Thermodynamics, 1154
- Three-month lag effect, 603, 604
- Three-stage least squares (3SLS) method, 2615
- Time-based incentive plan, 2147
- Time priority, 1095
 - rule, 998
- Time-series momentum (TSMOM), 1111–1114
- Time-series tests
 - factor model, 1468–1469
 - individual *t*-test, 1469
 - joint *F*-tests, 1469–1470
- Time-varying conditional risk exposures, 1116
- Tobin's *Q*, 1686, 1689, 1696, 1697, 2601, 2603, 2606, 2609, 2611, 2613, 2614, 2617
- Tobin tax, 1532
- Tokyo stock exchange (TSE), 1012, 1095

- Toronto Stock Exchange (TSX), 1095
 - CATS, 1096
 - floor-based system, 1095
- Total accruals (TACC), 1851, 2505
- Total assets, 1987, 1988, 1994
- Total return swaps, 577–578
- Total shareholders return (TSR), 2495
- Tracking-at-Risk (TaR), 2290
- Tracking error, 2288
- Tracking stock, 2472
- Trade credit, 2215
 - provision, 2220–2225
- Tradeoff hypothesis, 704
- Trading costs, 1004
- Trading mechanism, 1032
- Trading rules, 940
- Trading strategy, 2178, 2180, 2181, 2190, 2200, 2209
 - tail risk as risk measure, 2182
 - tail risk protection strategy, 2183–2185
- Trading volume, 1117
- Traditional agency theory, 793
- Traditional brokerage services, 755
- Traditional M&As, 2459
- Traditional/standard exchange rate
 - determination models, 946
- Transaction(s), 949
 - demand, 1132, 1134
 - explanation, 732
 - costs, 952, 955, 956, 958, 960
 - tax, 1532–1534, 1536
- Transaction costs, 1165, 1533–1535
 - of refinancing and default, 1175
- Transfer pricing (TP)
 - activities, 1542
 - actual variable cost, 1550, 1551
 - cost based, 1548
 - divisional managers, 1546
 - economic models, 1554
 - financial control systems, 1542, 1543, 1559
 - full cost transfer prices, 1551
 - fundamental issue, 1547
 - idle capacity, 1556, 1557
 - income tax burden, 1542
 - international transfer pricing, 1557–1559
 - linear programming model, 1554, 1555
 - management, 1542
 - management control systems, 1543
 - managers and other employees, 1542
 - market based transfer price, 1548
 - models, 1554
 - multi divisional firms, 1542
 - negotiated, 1552, 1553
 - no idle capacity, 1557
 - optimal production, 1544
 - selling division, 1542
 - shapely value, 1555
 - standard variable cost, 1549
 - variable cost, 1548, 1549
- Transitory dividends, 1603
- Transitory income, 1602
- Transmeta, 903
- Trans Union case (1985), 2480
- Treasury bond futures options, 820
- Treasury bond market, 610
- Treasury bond options, 820
- Treasury Inflation-Protected Securities (TIPS), 699
 - in 1997, 602, 610
 - as an asset class, 605
 - conversion from real to nominal prices, 603
 - cost, 606, 607
 - disadvantage, 605
 - inflation information aggregation, 608, 609
 - inflation risk premium, 606
 - information content of maturing, 608
 - information risk, 609, 610
 - liquidity premiums, 607
 - market size, 602
 - observable expected real rate, 607, 608
 - price discovery, 609, 610
 - price index, 602
 - public issuance, 604, 605
 - “real” rate of return, 602
 - reference CPI, 603
 - three-month lag effect, 603, 604
- Treasury bond market, 610
- Treasury’s Office of Debt Management (ODM), 606
- Treynor, J.L., 1401
- Treynor index, 763, 767
- Treynor-Mazuy model, 1280
- Treynor measurement, 989
- Treynor ratio, 986, 1399, 1401, 1402, 1404, 1407
 - and Jensen’s alpha, 2629
- Treynor Squared, 985
 - measurement, 989, 990
- Triangular fuzzy numbers, 2663, 2664, 2668, 2671
- Triggering event, 2478
- Trivariate dynamic conditional correlation model, 2583
- Trivariate structural relationship, 1621
- Twin security, 897, 898, 902, 907, 910, 920, 922, 926–929, 932, 933

- Two-bond portfolio, 687
- “Two-dimensional” thinking, 899
- Two-factor duration model, 689, 690
- Two-factor Markov process, 1302
- Two-factor structure model, 1572
- Two-index market model, 1183
- Two-stage compound option, 904–907
- Two-stage least squares (2SLS) analysis, 794
- Two-stage least squares (2SLS) estimation, 707
- Two-step acquisition, 2469
- Two-tier Corporate Board, 1191
- Two-tier Corporate Governance Structure, 1191
 - Chinese firms, 1192
- Type I and type II errors, 779, 785, 787
- U**
- Unbiased forward rate hypothesis, 598
- Uncertainty, 896, 898–901, 903, 904, 922, 2656, 2657, 2659, 2660, 2664, 2669, 2674
- Unconditional distribution, 2637
- Unconditional expectation, 630
- Underfunded pension liabilities, 2044
- Underinvestment problem, 967
- Underreaction, 1114
- Undervaluation, 1062–1064
- Unemployed CEOs, 2141, 2164, 2170
- Unfunded pension liabilities, 2052
- Unfunded scheme, 524
- Unified model, 1487, 1494
 - benchmark securities and model calibration, 1503
 - to financial crisis 2008, 1502–1503
 - functional engineering, 1505–1506
 - implications to risk management, 1502
 - interest rate risk, credit risk, and liquidity risk, 1504–1505
 - stochastic process of, 1499
- Union workers, 1143, 1144
- Unitary Corporate Board, 1191
- Unitary Corporate Governance Structure, 1191
 - functions, 1191
 - members, 1191
- Unitary corporate structure, 1192
- United States (US), 1225
- Unit labor cost, 1143, 1144
- Unit root tests, of wavelet components, 2583
- Univariate EGARCH model, 2583
- Univariate GARCH model, 2583
- Univariate tests, 802, 804
- Unocal anti-takeover defenses, 2480
- Unocal vs. Mesa*, 735
- Unqualified radical functions, 1950
- Unqualified utility functions, 1949–1950
- Upper Echelons Theory, 1689
- U.S. antitrust laws, 2482
- USA pure earnings yield factor portfolios, 2307
- USA pure momentum factor portfolios, 2308
- U.S. bank creditors, 1183
- US banks
 - assets, 1995
 - bankruptcy risks, 2002
 - bond index, 1995, 1996, 2010
 - cash, 1997, 1998
 - extra-ordinary risk, 2002
 - federal funds sold and securities, 1997, 1999
 - interest rate, 1995
 - L/A ratio, 1996
 - loans and leases, 1997, 2000
 - market index, 1995, 1996, 2010
 - monthly parameters, 2002
 - normal distribution, 2001, 2004
 - risk-based capital requirement, 1993
 - risk-free rate, 1995
 - stock prices, 1997
 - treasury’s bank bailout, 1996
- US brewing industry, 884
- U.S. Congressional debate, 1178
- U.S. Federal Government (SEC), 2460
- U.S. Internal Revenue Code, 1558
- U.S. Internal Revenue Service (IRS), 1558
- U.S. Merger waves, 2460, 2462, 2495
- U.S. nonfinancial corporations, 1654
- U.S. public-traded firms, 794
- US Securities Act of 1933, 968
- U.S. Securities Acts Amendments, 1016
- U.S. STRIPS bonds, 821
- Utility function, 1585, 1595, 2636, 2638
 - comparison of, 1958–1960
 - with favorable outcomes, 1952
 - with unfavorable outcomes, 1954–1956
- V**
- Valuation, 1642–1644, 1646, 1648–1652, 1654, 1655, 2460
 - model, 754, 756, 757
 - uncertainty, 1118
- Value Additivity Principle, 880, 885
- Value-at-risk (VaR), 750, 1364, 1443–1445, 1882, 2178, 2179, 2181–2184, 2193, 2200, 2209, 2558, 2576
 - backtesting analysis, 2591
 - conditional coverage, 2577–2578
 - daily forecasts, 2579
 - diagnostic test, 2567
 - dynamic quantile test, 2578–2581

- market risk measurement based on,
 - 1883–1884
- models, 1984, 1985, 1987, 1989–1991, 1993
- unconditional coverage, 2577
- Value relevance, 2046, 2047
- Variance decomposition, 2244, 2246
- Variance inflation factor (VIF), 2097
- Variance ratio, 1003
- Varspread strategy, 2180, 2181, 2200, 2203,
 - 2205, 2208
- Vasicek model, 826, 829, 832, 1739
- VECH model, 2558
- Vector autoregression (VAR), 2564, 2565, 2567
 - model, 926–928, 933, 949
- Vector error correction (VEC) model, 1142
 - autoregression, 2240, 2242
 - cointegration equation, 2239
 - commercial and industrial loans, 2238,
 - 2239, 2241
 - degree of endogeneity, 2242
 - dependent variables, 2242
 - error-correction, 2240
 - Hannan-Quinn information criterion, 2237
 - impulse responses, 2244, 2245
 - interactive term analysis, 2239
 - Johansen co-integration test, 2238
 - likelihood-ratio test, 2240
 - loan size and tightened loan standard, 2236
 - loan size and tightening banks, 2240
 - Schwarz information criterion, 2237
 - short-run and long-run granger causality,
 - 2243
 - speed-of-adjustment coefficient, 2242
 - unit-root test, 2236
 - variables, 2242
 - variance decomposition, 2244, 2246
- Venture capital business, 904
- Vertical mergers, 2460
- Veterans Administration (VA), 844, 845
- Virtual organization (VO), 1874
- Volatility, 721, 744, 747, 920, 1018, 1020,
 - 1021, 1531–1533, 1535–1536,
 - 2178–2182, 2186, 2187, 2192,
 - 2194–2196, 2199, 2208
- forecast performance, 2576
- measure of cash flows, 1588
- transmission, 1412, 1413, 1420, 1424
- Volume, 1536
- Voluntary debt restructuring strategy, 889
- Voluntary foreign credit restraint (VFCR), 1225
- von Neumann-Morgenstern (vNM) utility
 - theory, 1940, 1941, 1947, 1949,
 - 1951, 1952
- W**
- Wage-growth rate, 1143
- Wage negotiators, 1141
- Wage-push inflation, 1138
- Wald tests, 2218
- Warren and Shelton (WS) model, 2345
- Washington-Beijing relationship, 1049
- Wavelet(s), 2581
 - coherence techniques, multiple and
 - partial, 2561
 - transform, 2580
 - volatility forecast, 2586
- Wavelet-based analysis, 2560
- Wavelet VaR (WVaR) estimates, 2561
- Wavelet VaR (WVaR) forecasts, 2587, 2588
- Weak-form tests, 940–941
- Wealth effects, 890
- Web-based brokers, 754
- Webstock(s), 756
 - frenzies, 755
- Weighted average cost of capital (WACC),
 - 1642, 1646–1648, 2485, 2486
- Weighted Average Life (WAL), 2269, 2274
- Weighted exogenous sampling maximum likelihood estimator (WESMLE),
 - 784
- Weighted probabilities, 1947, 1948
- Well-informed financier, 796
- Wells Fargo
 - capital ratios, 2003, 2006, 2012
 - estimation, 1996
- Western creditor banks, 1178
- Weyerhaeuser Co, 1608
- White knight, 2478
- Wholesale Price Index (WPI), 602
- Wiener process, 1074, 1075, 1084, 1166, 1734
- Winbond Electronics Corp., 926
- Wooldridge's test, 1822
- Working capital (WC), 652–653, 1851
 - components, 655–656
 - corporate working capital, 653–655
 - decision, 2344
 - management, 653
 - net working capital, 653
- Working capital accruals (WCA), 1855, 2076
- WRDS European Short Data, 2326
- Y**
- Yahoo, 756
- Yale Model, 1254
- Yu, S.M., 1398
- Yuan internationalization, 1050

Z

- Zero-beta CAPM, 748
- Zero-beta portfolio, 748
- Zero coefficients, 1181
- Zero-coupon bonds, 542, 546, 683, 687
 - corporate and municipal, 543
 - face value, 546
 - financial market, 546
 - implicit coupons, 543
 - junk bonds, 543
 - Macaulay duration, 544–547
 - municipalities, 543
 - 5 or 10 year maturity, 546
 - pure discount bonds, 542
 - value loss, 544–546
- Zero-coupon Treasury securities, 542
- Zero earnings, 2062
- Zero intelligence agents, 1030
- Zero-intercept assumptions, 779
- Zulkaffi, A.H., 1398